PARALLEL CORPUS MULTI STREAM QUESTION ANSWERING WITH APPLICATIONS TO THE QU'RAN

AISHA JILANI

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Huddersfield

Submission: June 2013

The knowledge from which no benefit is derived is like a treasure from which no charity is bestowed...

Prophet Muhammad ﷺ

Book Reference: Saying of Muhammad, Compiled by Sir Abdullah Suharwardy

Abstract

Question-Answering (QA) is an important research area, which is concerned with developing an automated process that answers questions posed by humans in a natural language. QA is a shared task for the Information Retrieval (IR), Information Extraction (IE), and Natural Language Processing communities (NLP). A technical review of different QA system models and methodologies reveals that a typical QA system consists of different components to accept a natural language question from a user and deliver its answer(s) back to the user. Existing systems have been usually aimed at structured/ unstructured data collected from everyday English text, i.e. text collected from television programmes, news wires, conversations, novels and other similar genres. Despite all up-to-date research in the subject area, a notable fact is that none of the existing QA Systems has been tested on a Parallel Corpus of religious text with the aim of question answering. Religious text has peculiar characteristics and features which make it more challenging for traditional QA methods than other kinds of text.

This thesis proposes *PARMS* (**Par**allel Corpus **M**ulti **S**tream) Methodology; a novel method applying existing advanced IR (Information Retrieval) techniques, and combining them with NLP (Natural Language Processing) methods and additional semantic knowledge to implement QA (Question Answering) for a parallel corpus. A parallel Corpus involves use of multiple forms of the same corpus where each form differs from others in a certain aspect, e.g. translations of a scripture from one language to another by different translators. Additional semantic knowledge can be referred as a *stream* of information related to a corpus. PARMS uses Multiple Streams of semantic knowledge including a general ontology (*WordNet*) and domain-specific ontologies (*QurTerms, QurAna, QurSim*). This additional knowledge has been used in embedded form for *Query Expansion*, *Corpus Enrichment* and *Answer Ranking*.

The PARMS Methodology has wider applications. This thesis applies it to the Quran – the core text of Islam; as a first case study. The PARMS Method uses parallel corpus comprising ten different English translations of the Quran. An individual Quranic verse is treated as an answer to questions asked in a natural language, English. This thesis also implements PARMS QA Application as a proof of concept for the PARMS methodology. The PARMS Methodology aims to evaluate the range of semantic knowledge streams separately and in combination; and also to evaluate alternative subsets of the DATA source: QA from one stream vs. parallel corpus. Results show that use of Parallel Corpus and Multiple Streams of semantic knowledge have obvious advantages. To the best of my knowledge, this method is developed for the first time and it is expected to be a benchmark for further research area.

Table of Contents

(word count: approximately 48,000 words)

List of Figures

List of Tables

Dedicated to my husband, Shakeel Ahmed

The wind beneath my wings

Acknowledgements

"Do not consider even the smallest good deed as insignificant; even meeting your brother with a cheerful face (is a good deed)." *Prophet Mohammed* ﷺ

This piece of text might read more like a story instead of formal words. This is the best way I can express my gratitude to those whom I owe great respect and appreciation. This is because I see my PhD as a long journey not only a degree. I also want to share my personal experience with those who are held back thinking that a PhD has no end. I have to acknowledge many people; as I met so many 'lovely' personalities on my PhD journey who have contributed in my success one way or the other.

I see this research work as produced by a first time mother (myself) who knew nothing BUT had the will to achieve a PhD besides having a lovely & lively family. Honestly, every time I felt low my considerate supervisor restored my moral by sharing his personal experiences. He kept me reminded how much he trusted my skills. He made me realise that a PhD is not about earning a degree; it's about learning to remain determined and be productive despite how *low-flight* the journey is or how high your research target is. I was given more *hopes* by a kind-hearted lady. She gave me the directions which led me to a point where I was free to think and pave my way to the end. But there still was something important missing – the test data. Here another sympathetic and intellectually generous personality allowed access to his expertise. He encouraged me to identify and fit in the missing pieces of the jigsaw. His dedicated support and encouragement did wonders and I achieved the amount of work in minimum time which I'm sure must have been very difficult otherwise. Finally, following are the names of the respective personalities:

- Prof Lee McCluskey: Thank you Lee for having faith in my abilities and helping me make my dream come true. My sincere respect to you for being my supportive supervisor 2nd time.
- Dr Di Cai: Thanks a lot for showing me the direction where the solution to my problems laid.
- Dr Eric Atwell: Words are not enough to say what I got from you when and where I needed it. Thanks a lot for being around.

I would also like to thank the PARK group members; Dr Diane Kitchin and Dr Andrew Crampton for the feedback on my 'lengthy' presentations. I would like to mention the critical thoughts I got from Dr Rashid Mahmood; Sir your support kept me on track in my confused times. Dr Julie Wilkinson I have always adored you for your moral and personal support in all matters. Thanks for everything Julie. Dr Zhijie Xu you will always be remembered in good words for your kindness and support. I must thank Ms Gwen Wood

and Chris Sentence for taking care of all admin issues that have arisen in this time. Thanks a lot IO Team, Dr Ian Pitchford and Gail Hurst for helping with related matters.

It is important to name my friendly colleagues who made our office environment *'less freaky'*. Rabia Jilani (my sister) thanks for being *there and everywhere* in my final year and yes! also for decorating my desk every now and then to cheer me up. I won't let you miss me...!!! Dr (to be soon) Shahin Shah I will miss our tea coffee and nibbling together. Falilat Jimoh, I'm grateful for the handy advice you always had for me. Asma Kilani, thanks for absorbing my free advice. Dr Fariha Chaudhary, thanks a lot dear for listening to my research issues and sharing your wisdom; you always gave me a different perspective for my work. Last thing – my dear office desk: I will miss you a lot too...!

Overall, my PhD journey was tough but all pains turned into gains as my child, Shaheer Ahmed, grew besides this research work. My husband, Shakeel Ahmed, ensured to suppress my guilt as being a working mother. Doing so, he took stress out of my life. He did everything and anything to give me the best opportunity to gain this success. Thank you so much Ahmed – for being '*true*'. I must have never achieved this without your support.

Now to mention contributions of my parents and family... Nothing in this world can be achieved unless you have your parents behind your back to trust on, to rely on, to lean on and to reflect on. I would like to express my deep gratitude to my parents, my father Prof. Ghulam Jilani and mother Shahida Jilani. Ammi and Puppa I am ME because of YOU. The more I am learning, the more I understand your unbeatable value in my life. Not to forget, Abba Gee, Maan Gee Juzak Allah Khair for your prayers and moral support. I also owe special thanks to my uncle Saeed Anjum for his critical thoughts on my work. I would also like to acknowledge his family for the on-going support they have been providing me throughout my life in UK. Last but not least, thanks a lot to my dear brothers Ali Asghar Jilani for lending me an ear whenever I needed one and Awais Jilani for the good laughs you always shared. This provided me a safe zone to relax at times. My sincere thanks to Anila Bashir for all the practical support I got from her.

An aspect of my work was to apply my research methods to the Quran. In this regard, I got several volunteer offers from several knowledgeable and motivated people across the globe. I have availed myself of a few of these and will surely get in touch whenever the need arises in future. I thank you all for your contributions in this regard!

Finally, I thank Allah SWT for enabling my 'Sudqa-e-Jariya' through this research work. May He bless me more success. Ameen!!!

Disclaimer

Applying my research to the Quran as a case study helped me better understand the topics and the Verses of this book. This is one of the best things I have achieved during my PhD. I have tried hard to accomplish this task in the best possible way. But I would also like to state that research presented in this thesis is a part of computer science studies, not religious studies. The method proposed in this work tries to compute answers from the Quranic translations using different question answering methods reputable in the subject area. Authors are by no means trying to supplant or override any human experts in this context. The answers provided by the proposed method should NOT be seen or used independently of the advice from domain scholars. Also the choice of English translations of the Quran was purely based on firstly, availability and then suitability for the task. We are not claiming any translation to be more virtuous than any other. May Allah SWT accept my limited effort – Ameen.

"He who makes peace between people by inventing good information or saying good things, is not a liar."

Prophet Muhammad ﷺ

Book Reference: Sahih Al-Bukhari, Volume 3, Hadith 857

Chapter 1 – Introduction

Advances in information technology and the volume of information available in electronic form led to the development of advanced search engines and document retrieval systems, e.g. Google and AltaVista. Although these systems provide links to relevant documents very efficiently, the user is required to take the trouble to search through them to identify the most suitable document which satisfies his information need. This is what triggered research into developing Question Answering (QA) Systems which were capable of mapping a user's knowledge requirement presented in a natural manner (i.e. by asking questions), to a piece of information that satisfies the user's needs (i.e. answer to asked question). Here the challenges are to determine the knowledge requirement precisely, identify where the required knowledge is available, and finally extract present knowledge to a level of user understanding and satisfaction. Development of a system capable of meeting these challenges has been a long-standing research goal.

Research on QA Systems has a history since the 1960's, resulting in multiple techniques, methods, procedures and several attempts to design successful systems. Question Answering (QA) is recognised as an inter-disciplinary research task that requires input from Artificial Intelligence (AI), Information Retrieval (IR), Information Extraction (IE), and Natural Language Processing communities (NLP). Evaluation campaigns like Text REtrieval Conference (TREC 1999-2007), Cross Lingual Evaluation Forum (CLEF), Question Answering track at Initiative for the Evaluation of XML retrieval (QA@INEX) and REsPubliQA 2010 have encouraged more research on QA systems design and development. Today several open-domain and closed-domain QA systems intended for a variety of purposes exist but their level of accuracy, efficiency and feasibility are still a matter of debate and require further research.

A major area of interest that needs the attention of Question-Answering community is the Religious Domain. Religious texts include the faith-defining religious canon, authoritative interpretations and commentaries, sermons, liturgy, prayers, religious poetry, and lyrics. There is computational research involving Religious Text e.g. Automated Classification, Cluster Analysis, Machine Learning, Named Entity Recognition, Collocation Extraction, Information Retrieval and Linguistic Analysis. But no prominent attempt to apply Question-Answering techniques to any religious texts was found by the author of this thesis.

Religious texts have been regarded as interesting and challenging for the language resources and evaluation research community (Atwell, Brierly et al. 2012). We believe

Religious texts are equally valuable for the Question Answering research community and any attempts in this direction are likely to attract more useful research. QA research involving Religious texts is bound to have a direct and positive impact on society by making the religious sources more accessible for the general public and further research. This is because there are billions of readers of the holy books. It is also known that followers of religion ought to refer to religious text very frequently for religious, personal and motivational reasons. At present human domain experts are mainly queried for information comprehension and question-answering. These scholars are available in person but since the evolution of the Internet they are now available through website forums also. But it is important to note that these services are closer to the definition of Information Dissemination systems. This is because they are partly automated and involve the human brain at the core level for answering user questions. Religious texts can prove not only an interesting but also challenging domain for the QA research community. This is because they inherit the traditional issues of natural language processing and also pose more peculiar challenges specific to the domain.

Rationale

We rationalize that Religious Question-Answering is recognisably different from factoid Question-Answering on generic text documents. It is generally believed that involved techniques are required to be adapted to the data characteristics (Pérez-Iglesias, Garrido et al. 2010). Hence a question arises whether methods and techniques that have proved effective in developing QA Systems to deal with traditional text sources are equally effective for Question-Answering with religious texts. If not, what techniques and tools need to be adopted? As per our knowledge, no QA system has previously been developed with the aim to attempt Question-Answering on religious texts. Hence this is a research question that needs examination.

A key aspect of research presented in this thesis is to investigate how to adapt generic Question Answering tools and techniques to the specialised needs of Religious Questions for finding answers in a parallel corpus of English translations of a religious text. The following are a few attributes that make the source text corpus ideal for the task:

- The text corpus should have no privacy constraints and costs involved. It should be available free for research purposes.
- A large community of experts for that corpus exists, which has developed 'linguistic tagging schemes' or ontologies for the domain.
- A large group of users exist, that can assist in evaluating the system and its results.
- A huge group of potential users exist, so the research in the domain has an impact

The Quran – the core text of the Islamic religion has been used as the first case study in this research. The Classical Arabic Quran meets the above listed criteria in their fullness as there are over a thousand of years of documented work on the linguistics of the Classical Arabic text of Quran. Besides most features and resources mentioned above are valid for the English translations of the Quran also. In today's global population Muslim readers of the Quran comprise "the largest user-group ever for a single text corpus" (Eric Atwell 2013). Out of this huge population, the audience user group for research on English translations of the Quran is still a huge group - the billions of Muslims worldwide who communicate using the English language. Quranic Arabic corpus linguistic research has produced ontological resources and tools such as the named – entity ontology (Dukes 2011), semantic concept "tags" (Abbas 2009), QurAna (Sharaf and Atwell 2012) and QurSim (Sharaf and Atwell 2012). Although these resources were developed for the Arabic Quran they remain equally useful for research on English translations of the Quran as the verse references are the same regardless of the language of the Quran. So they can prove useful resources for Question-Answering System design.

Motivation

Observing the Online Question Answering services generally used by the Muslim community, we have deduced that there is a big demand for being able to answer questions that need references from the Quran. Most Islamic websites providing QA service, e.g. Islam Question Answering (Al-Munajjid 2013), Study the Quran (Ali 2013), Islam City (Imam 2013), Quran Al-Islam (Salhey 2013), Islam Web (Mufti 2013), Qibla Answers (Keller 2013), On Islam (Scholars 2013), Understanding Islam (Amjad 2001), Alim (Shah 2010), Studying Islam (Rahim 2007), and Renaissance Journal (Sheikh 2006) rely on human scholars for answering user questions. Our interest is to investigate and evaluate to what extent Artificial Intelligence Computational methods can emulate or assist these experts. The QA System proposed in this thesis is potentially a decision support system not a system aiming at replacing the domain experts. Another aspect of the motivation is our personal interest in the Islamic domain in general and the Quran in specific.

Novel contributions of thesis

To the best of our knowledge, prior to this work, no fully-fledged computational PhD programme has been conducted for Question Answering (QA) on English translations of the Quran. The novelty of this research lies in attempting QA by applying existing advanced IR

techniques coupled with sophisticated NLP products, domain-specific ontological resources to novel domain of the Quranic text. Several advanced IR and NLP methods were required to be adapted for the purpose leading to new contributions within the domain. This work also puts forward the hypothesis that additional semantic knowledge adds value and helps in identifying correct answers. The novel contributions of this thesis to the subject area are as follows:

I – Question Answering System tailored to Religious Text
II – PARMS Methodology
III – Resource specific to QA of Religious Text – QurTerms
IV – Model Question Answer Set – PARMS Model QA Set
V – PARMS Evaluation tool

The work presented in this thesis is inspired by on-going recent research in this area that has been mentioned in the rationale. It is expected that contributions of this novel research will attract more researchers towards the subject. We intend to further extend this work and make the products of this research freely available for reuse under General Public Licence (GPL). A summary of the contributions of this thesis is given below.

I – Question Answering System tailored to Religious Text: QA research community experts indicate that Question Answering has benefited from collaborative work across related fields. They inform us that Question Answering depends mainly on two complementary fields; Information Retrieval (IR) and Natural Language Processing (NLP). Research in this thesis applies a combination of IR and NLP techniques for Question Answering and tailors this to suit the needs of a novel religious domain.

II – PARMS Methodology: This thesis proposes *PARMS* (**Par**allel Corpus **M**ulti **S**tream) Methodology; a novel method applying existing advanced IR (techniques combining them with NLP methods and additional semantic knowledge enrichment to implement QA (Question Answering) for a parallel corpus. Details of the method are provided in chapter 3.

III – Resource specific to QA of Religious Text – QurTerms: This thesis proposes a technique to devise a domain specific ontological resource. This technique has been applied to the Islamic domain as a sample but it has wider application across domains with similar characteristics. QurTerms resource is ontology of *Localised Quranic Terms & Concepts in English* derived from Quranic translations and commonly used Islamic Terms. The QurTerms ontology holds useful semantic knowledge of three basic types.

- Firstly, QurTerms provides different *linguistic representations* of the same Islamic terms as used by different authors, translators, scholars and general public.
- Secondly, QurTerms provides *meanings* for those Islamic terms which do not belong to contemporary English; hence their meanings are not available in glossaries and dictionaries.
- Lastly, QurTerms provides us *thesaurus* for the Islamic terms that are commonly used and take different shapes and forms depending on the context of their use. A thesaurus of these words is usually not available in the WordNet ontology.

QurTerms is an initial step contributing towards a solution to problems described later in this chapter, section 2.3.

IV – Model Question Answer Set – PARMS Model QA Set: The PARMS approach has been evaluated using a corpus of 100 Question Answer pairs called *PARMS Model QA Set*. This evaluation resource has been collected from different original resources. This work also attempts to analyse what kind of questions people are expected to ask a QA system in the Quranic domain. These model QA pairs use Natural Language, English in this case. These sample questions exemplify 'real questions' as they had been asked by 'real people' and answers are based on discussions and commentaries provided by 'real domain experts'. These have been designed for multiple purposes and have been gathered from multiple sources such as:

- Extracted from Authentic online scholarly sources
- Acquired from Educational Quranic QA websites
- Collected from Potential End users i.e. Domain Users
- Derived from all the above – Questions designed to meet specific experiments

Chapter 3 provides further details on the rationale, sources used and methodology adopted for gathering these sample questions and finally the design and representation of the PARMS Model QA Set.

V – PARMS Evaluation tool: The aim of the PARMS Evaluation tool is to test the PARMS Methodology in order to identify a combination of different parameters that produces the best results. The evaluation has been carried out to test the performance of domain-independent methods i.e. without using domain-specific resources and compare it to the results obtained using domain-specific knowledge with application to the Quranic dataset to test the hypothesis behind this thesis. The author has developed the PARMS evaluation tool, as a proof of concept for the PARMS methodology which is specifically tailored and adapted to cope with the specific challenges of religious text genre.

PARMS evaluation tool is not an end-user ready QA system. Instead it is just a tool to evaluate PARMS methodology. This is because it is configurable to evaluate the effect of different parameters on PARMS methodology. The PARMS Methodology aims to evaluate the range of semantic knowledge streams separately and in combination; and also to evaluate alternative subsets of the DATA source: QA from one stream vs. parallel corpus. Details of parameters used for evaluation and available options have been provided in Chapter 4 section 4.1.

Thesis structure and highlights

This section is a discussion how the remainder of the thesis is structured. *Chapter 2, Literature Review,* presents general subject areas around this thesis including following:

2.1 Theoretical Underpinnings

- Natural Language (NL)
- Question Answering (QA)

2.2 The Quran Corpus
2.3 Computational Research on the Quran

Chapter 2, section 2.1 reflects on how natural language is challenging and ambiguous and what problems an automated Natural Language Processing (NLP) would be expected to deal with. Moving on, chapter 2 reports on evolution, evaluation campaigns and general architecture of Question Answering (QA) systems. It introduces a few traditional QA systems and multi-stream QA systems that provide the baseline for this thesis. A critical analysis of QA systems is presented towards the end.

Later *section 2.2 – The Quran Corpus,* introduces readers to the religious text corpus used as a case study and data set for this research. It starts with some background information on the Quran, its structure and language and prominent topics of interest in connection to Question Answering on the Quran. The Quranic text has peculiar characteristics which have been presented with examples in a dedicated sub-section of this chapter. Section 2.2 also introduces the ten translations used for this thesis. Why Question Answering is required on the Quran and what makes it difficult has been discussed in the next section of this chapter. Lastly there are a few indicators of literature where other religious books have been considered for computational research on religious texts.

The last *section 2.3 – Computational Research on the Quran,* presents recent research where the Quran has been involved as a case study. It covers a variety of research in this area including Cluster Analysis, Automated Classification, Visualisation Systems, Ontologies, Information Dissemination Systems, Verse Retrieval Problem, Corpus Development and Linguistics. The chapter also lists the research work i.e. development of *QurAna* and *QurSim,* which have had a strong impact in convincing me to go ahead with the research that has culminated in this thesis. Besides published research some in-progress research (which has been happening in parallel to this thesis) is also included in this section. Hence these projects are either incomplete to present or have limited publications but they have been included for the sake of completeness. Another motive behind compiling this section is to provide a one-stop source of up-to-date research involving the Quran which is not readily available otherwise. Lastly the chapter also introduces *QurTerms Ontology*, which is a contribution of the present research work as listed under contributions of this thesis in the previous section.

The fundamental research behind this thesis is detailed in *Chapter 3, PARMS*. The chapter introduces the core methodology in much detail supported by diagrams and other visual aids to support readers' understanding. The chapter briefly describes the theoretical underpinnings that lay the foundations of the proposed methodology. The methodology comprises Phases I – V which are elaborated in simple words for maximum readability. Next the chapter describes the implementation details of the application developed as proof of the concept of this research. Finally the chapter also provides a visual walkthrough of the system designed for evaluating the proposed methodology. The evaluation methods adopted are discussed in the next chapter.

Chapter 4, PARMS Evaluation System, as the name suggests, provides a detailed description of the evaluation system prior to actual evaluation results. It details the major components and overall framework layout, i.e. set of evaluation questions and evaluation measures adopted. This chapter firstly introduces another contribution of this thesis, the *PARMS Model Question Set*. This set of question-answer pairs was compiled in connection to the evaluation of the proposed methodology. The description includes the rationale behind this source, collection sources used, classification of Q&A pairs, statistical information about the set and technical representation used for compilation. The chapter proceeds with a manual-type step-by-step description of each parameter with its purpose and sub-options that has been used by the evaluation system. Evaluation groups, metrics used for evaluation and the visual aids used to describe the results have been reported towards the end of the chapter. Two new visual aids devised are also introduced, namely the Bitmap Image table and Change Chart diagram. These were required as we were unable to identify any existing

visual aid suitable to convey certain aspects of the results more efficiently. Knowledge of all contents discussed in this chapter is required in order to better understand the evaluation results presented in the next chapter.

Chapter 5, PARMS Evaluation Results, reports the evaluation of the results for the proposed methods, techniques and choice of resources. Experiments have been grouped on certain parameters for effective comparisons. Trends identified for each group of results have been presented using visual aids for better understanding. Initially some preliminary experimental results have been discussed which led to the detailed experimentation reported later in the this chapter. Finally analysis of the results has been presented using random examples of experiments to highlight the performance of the proposed method. Generally experiments are described in a walkthrough manner where visual elements like screen shots, graphs and tables are used to facilitate understanding. Analysis and discussion on each of the result set has been provided along each example and experiment.

Finally the thesis closes describing the conclusion of this research in *Chapter 6 – Conclusion & Future Work*. The same chapter proposes some future work directions. This includes our comments on how the PARMS QA System can benefit from further research in certain dimensions and what other techniques can be explored to address the same research problem. It also informs the reader about what impact this research can have on the subject area and how certain contributions of this thesis can be reused for similar domains.

Summary

This chapter introduces the background of this research, its rationale and the motivation behind the work presented in this thesis. It also briefly informs the reader about the novel contributions of this thesis. Finally it provides an outline of the thesis structure giving highlights for each chapter.

Chapter 2 – Background Studies

2.1 Theoretical Underpinnings

This research is an inter-disciplinary project and proposes a methodology of Question Answering benefiting from existing advanced Information Retrieval techniques coupled with sophisticated Natural Language Processing (NLP) techniques using domain-specific resources.

Natural Language Processing (NLP) as a field of computer science is classed as a sub-field of Artificial Intelligence (AI) due to the nature of the problems (complexity and ambiguity) it poses. Mufti (2013) cites H.G.Wells' book, The Time Machine (published 1895, reissued 1995) as an example and states that "the interaction in natural language form has been a dream of artificial intelligence since the invention of computers and even foreseen earlier by creative imagination". NLP includes the core technical methods and theories that surround manipulation of human natural language using a computer machine. Roger Schank and Alex Kass suggest that in order to make progress in NLP and design intelligent machines that understand natural language as human beings do, researchers must address such questions as how we (human beings) understand and represent the concepts that language can communicate, how we learn new concepts, and how we organize this knowledge in memory.

This chapter is organised as a literature review of subject areas that have interplay in this research. The chapter starts with describing issues around Natural Language Engineering (NLE) and Natural Language Processing (NLP). Question Answering (QA) Systems are discussed in much more detail. The discussion provides a quick walkthrough of the evolution of the early and recent QA systems. Then a critical review of QA systems in connection to their application to religious datasets is presented. Discussion follows on describing the idea behind multi-stream QA and the final section informs readers about NLP techniques involved in this research work.

I – Natural Language

Shuly Wintner in his research article – *what science underlies natural language engineering (NLE)* comments that "What makes our (NLE) systems special is the fact that they manipulate natural languages (e.g. English, Arabic, etc.), and the only scientific field that can inform our work is **linguistics**" (ISOC 1980). Wintner also insists that "Not only should

we be more aware of linguistic research that can improve our (researchers in NL) engineering work, we (researchers in NL) should also be directly involved in such research" (Wintner, 2009) that can inform design of systems that deal with natural language in a more formal way.

Natural Language – Challenging

Formal computer programming languages such as C++ or Java are strictly guided by the grammar rules and violation of the simplest rule can generate an error. Spoken language also has set rules to be followed yet minor syntactic errors do not deteriorate the semantics of the communication. NIST (1992) argues that this flexibility of natural language is what is difficult to achieve with language understanding computer systems because "What is communicated by natural language is not explicitly stated" (Scholars 2013). Natural Language researchers agree that at an ultimate level of natural language understanding, actually we need to understand the syntax, the semantics and pragmatics of the sentence. Syntax is defined by the grammar of the language. If a sentence is syntactically well formed it helps in understanding the meaning of the sentence. But it is also important to note that sentences that are syntactically correct are not essentially correct in their meaning. For example the sentence "*Colourless green ideas sleep furiously"* is grammatically correct but it is still meaningless.

Edward Sapir, the famous 20th-century anthropological linguist, presents juxtaposition for his celebrated sentence *"The farmer kills the duckling"* as *"The man takes the chick"*. He claims that the new (latter) sentence is totally different from the first sentence in what it conveys, not in how it conveys it (Ali 2003); even when apparently both the sentences appear to fit the same syntactic pattern. Elaborating on Sapir's work the Russian linguist Roman Jakobson argues that if the above sentence is conversed and codified in some purely relational and non-material regard, and (for instance) changing the word order ***A*** *kills* ***B*** by the inverse sequence ***B*** *kills* ***A***, we do not vary the material concepts involved but (only change) uniquely their mutual relationship (Jakobson, 1980). Although this change of relationship does not affect the syntax of the sentence, the sentence now does not *'make sense'*. As agreed both semantics and pragmatics play a vital role in understanding the sentence uttered in natural language.

NIST (1992) believes that common sense knowledge, reasoning capability and experience with the use of language helps a human being interpret meaning. The famous corpus linguist Wallace Chafe comments, "Just what does it mean to understand something? I believe that understanding depends on the ability to place ordinary, particular experiences,

whether they are derived through the senses or through events within ourselves, within some larger picture where they *make sense*..." (Marican 2013).

English Language – Ambiguity

Edward Sapir declares that "It would be impossible for any language to express every concrete idea by an independent word or radical element. The concreteness of experience is infinite; the resources of the richest language are strictly limited" (Ali 2003). The inherent ambiguity in English language poses several problems for Natural Language Processing (NLP) systems. Some words in the language do not have a single meaning. To *make sense* of the meaning, the context, the word lies in needs to be understood e.g.

We gave the monkeys the bananas because ***they*** were hungry
We gave the monkeys the bananas because ***they*** were over-ripe

In the above sentence the referent of the use of word *'they'* is ambiguous. The true meaning can be identified only if the properties of the objects – monkeys and bananas are known.

English does not also distinguish the relationships between parts of speech (e.g. which noun a particular adjective refers to) in some sentences. For example in the phrase *'Pretty **little** girls' school'* it is hard to tell which word the adjective *little* applies to. This could apply to either the *girls'* or the *school*. It could be taken as to be connected to the word *pretty* as well.

Information also adds meaning to the overall sentence. For example in the sentence *'I never said she stole my money'* different emphasis on each word changes the entire meaning of the sentence.

```
sentence      -->   noun_phrase, verb_phrase.
noun_phrase   -->   determiner, adjective, noun.
noun_phrase   -->   adjective, noun.
noun_phrase   -->   determiner, noun.
noun_phrase   -->   noun.
verb_phrase   -->   verb, noun_phrase.
verb_phrase   -->   verb, preposition, noun_phrase.
determiner    -->   a | an.
adjective     -->   fruit.
noun          -->   flies | fruit | time | arrow.
noun          -->   banana.
verb          -->   like | flies.
preposition   -->   like
```

Figure 2.1: Grammar for the sentence 'Fruit flies like a banana'

Technically there could be multiple parse trees for a single sentence which again needs deciding on which one to use. For instance for the sentence *'Fruit flies like a banana',* two possible parse trees shown in figure 2.2 and figure 2.3, can be derived from the grammar in Figure 2.1.

The first possible parse tree can be derived as follows:

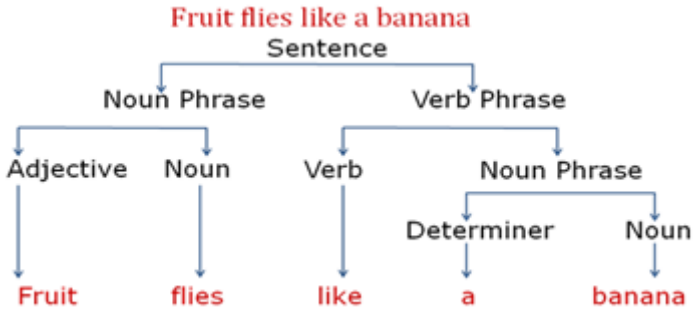


Figure 2.2: First Possible Parse Tree

Another possible parse tree for the same sentence and from the same grammar can be as:

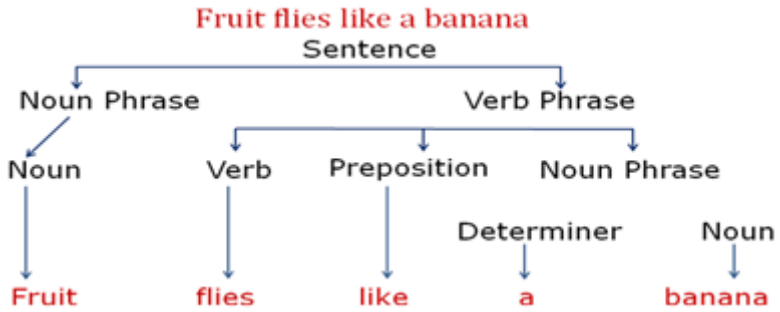


Figure 2.3: Second Possible Parse Tree

The sentence above also has a problem that it can be considered to have multiple meanings at the same time. For example there can be several interpretations for the sentences given below:

- Time *flies* like an arrow
- Fruit f*lies* like a banana
- He is my *old* friend
- My brother Sunny grew *another foot*

It is also important to understand that the appearance and structure of the word changes in different situations. For example in the plural form *child* becomes *children* whereas *book* becomes *books*. To achieve an ideal NLP system such challenges need to be addressed.

Dealing with Ambiguity in Natural Language

Natural language processing (NLP) has long been confronted with the problem of dealing with challenges arising due to the ambiguous nature of natural language. Church et al. (Amjad 2001) believe that natural language sentences can be far more ambiguous than one might think. They indicate that sentences written in a natural language can sometimes produce thousands of syntactic parse trees resulting from the ambiguity in interpretation of the meaning of certain clauses in the sentence. Church makes an observation that a sentence with only two prepositional phrases (PP) such as, *Put the block in the box on the table* - has two interpretations as follows:

- Put the block [in the box on the table]
- Put [the block in the box] on the table

But in fact, this number of interpretations 'combinatorially' grows. The above sentence bears five interpretations with an additional prepositional phrase, as in:

- Put the block in the box on the table in the kitchen

The reader can verify for himself that there would be fourteen interpretations with four PPs. Each of these interpretations bears a different binary tree. Church suggests analysing this structural ambiguity as the product of its components or as a series of connections.

Schelenker (2006) argues that sentences are sometimes assumed to be mere strings of words put together, which actually is not the case. They are actually made of word(s) that *form constituents* structured in a *tree-like* fashion. Hence sentences are represented as *syntactic trees* in linguistics. The author justifies this argument with a *constituency test* that makes use of the juxtaposition of words. The author notes that ambiguity in a sentence can be *lexical ambiguity* caused by the presence of an ambiguous word e.g.

- He was sitting beside the *bank*

In the above sentence word *bank* can refer to either a financial institute or the river side. Another type of ambiguity is *structural ambiguity* leading to generate more than one syntactic tree for the same sentences but this can be resolved. The author makes an interesting prediction after empirical analysis, "A well-chosen constituency test should disambiguate the sentence".

II – Question Answering Systems

Today's search engines like Yahoo, Google and MSN have made information access a lot easier. These perform keyword search using techniques such as link analysis and counting number of words in the query and this has enhanced their capability to *search* information from most popular and lexically related pages. However, search engines are not designed to deal with natural-language questions (Keller 2013). An emerging alternative to keyword-based web searching is automated question answering (Mufti 2013). A simple experiment conducted by Roussinov et al (2008) proves that the QA ability claimed by Google and Microsoft search engines is as yet limited to certain areas (Google - questions related to geographical data, MSN – questions based on Encyclopaedia Encarta) as they can sometimes produce useless and even false results for questions on dynamic topics ("Who is the CEO of ..."). They both lack the ability to answer semantically complex and challenging questions ("How long can a British Prime Minister serve in office?"). These shortcomings highlight need for a more efficient way to deal with information requests placed in natural language. This has led to focused research on *Question Answering Systems* within Natural Language processing research.

Question Answering (QA) systems can be defined as automated systems capable of processing a question asked in natural language, e.g. English or Arabic and returning the *"Answer"* providing the required information. In-depth analysis informs us that an *Ideal* QA system should be able to understand the required information through simple natural language, locate this required information, extract this information from its source and return it in the form of an answer.

Although there has been decades of research aimed to achieve this *ideal* QA system, the state of the art is still far away from achieving these goals. To achieve this vision, the QA research community has divided the overall QA task into specialised sub-tasks, e.g. Information Retrieval (IR) to feed the information requirement and Natural Language Interfacing using Databases to support information retrieval from structured information sources. This has helped concentrate efforts on more manageable problems which can also inform other problems in parallel such as Question Answering.

Later Question answering (QA) received attention from the information retrieval, information extraction, machine learning, and natural language processing communities (AAAI, 2002; ACL-ECL, 2002; Voorhees and Harman, 2000, 2001, cited by (Jones 2004)). By acquiring AskJeeves.com (now called Ask.com) in July 2005, InterActiveCorp wanted to achieve a completely open-domain question answering system that (can) prove to be the holy grail of information access (Keller 2013).

Evolution of QA

At the early stage of NLP, understanding the meaning of a sentence, dealing with ambiguity, translating into formal languages and searching large tree structures were seen as the key problems in NLP. These were also seen to propagate within application of NLP e.g. QA. From the very early stage of QA systems, they were seen from two perspectives: Artificial Intelligence (AI) and Information Retrieval (IR).

Simmons (Aznan Zuhid Saidin 2005) indicates that one of the early written programmes, Belnap logically analyse a question in an attempt to provide an answer. The classification breaks the questions into two parts, i.e. one part that delineates a set of alternatives and another part that makes the request. These answers were classified as direct answers, partial answers, eliminative answers, corrective answers and relevant answers. Other approaches classified questions into complete or incomplete depending on the kind of information they are expecting.

The history of programs written to answer questions goes back to the early 1960's in the form of the Natural Language Database QA (NLD-QA). There was a motivation to find an alternative to formal query languages. This motivation was derived from the AI influence on QA Systems, where knowledge was encoded using structured databases. NLD-QA approach was to translate the question into a formal database query, find the answer and translate the results back into the question language. The advantage of this approach was that as the conceptual schema of the application domain was known, it was possible to apply deep reasoning and other modern techniques. Initial attempts in the area were programs like, *The Conversation Machine.* This was designed with the intention to gain experience with the meaning of *Turing's* idea. It did not attempt to tackle the core NLP problem, i.e. syntactic analysis. Instead it dealt with it using the principle of an attribute-value pair. Hence this program was limited to dealing with simple English sentences alone.

Another experimental system, *The Oracle*, developed using LISP, made use of syntactic analysis by structural matching of syntactic-semantic codes. This system dealt with questions that comprised only one subject and/or object and could not go beyond this. The sentence Appraiser and Diagrammer and Semantic Analysing Machine (SAD SAM - 1963) system was capable of handling Basic English (a certain subset of English) sentences. It comprised different sections for parsing using predictive-analysis techniques and handling semantics. In contrast to NLD-QA, the aim of a text-based QA system is to attain the capability to answer a question narrated in natural language. Protosynthex (1963) made use of indexing, synonym dictionary, intersection logic and information scoring to deal with questions.

This trend shifted to Question Answering over free text in the mid-1990's. Dalamas (Mohd Yunus 2008) indicates that there was more work on Reading Comprehension (RC) and open domain QA during this period. Both made use of Boolean query and/or regular expressions to approach the problem of QA. This involved use of an Information Retrieval (IR) module that extracts the candidate answer(s) from the relevant text or document. Each answer was ranked for its occurrence and the final answer calculated. This change in approach led to certain challenges such as ambiguity recognition and resolution. In RC and open domain QA is delayed until after the data retrieval stage whereas this is dealt with before database access in NLD-QA. This is due to the tight coupling of question and answer in the latter as compared to the former.

Researchers from the early period of question answer systems convincingly identified that "generalizing from current small–scale experiments to language-processing systems based on dictionaries with thousands of entries – with correspondingly large grammars and semantic systems – may entail a new order of complexity and require the invention and development of entirely different approaches to semantic analysis and question answering" (Shah 2010).

QA Evaluation Campaigns

Although current QA evaluation forums TREC, CLEF and NTCIR are termed competitions they are more correctly thought of as Evaluation Campaigns. They are focused on lab-to-product transfer of text retrieval technology. Each conference culminates with a workshop where the participating systems share their learning and understanding of the domain.

In late 1990's the perspective on QA System was heavily influenced by TREC and similar evaluation campaigns. Text REtrieval Conference (TREC) was the first evaluation campaign started in 1992 by NIST. The initial aim was to set out a benchmark for ad-hoc search and text document retrieval. The first QA evaluation took place in 1999 (Hammo, Abu-Salem et al. 2002). This forum explored different aspects of QA over the years. TREC-8 QA participants were provided with 200 questions and the participating systems were expected to offer an answer of 50 – 250 bytes length. TREC-9 QA aimed at studying the impact of formulating questions in different ways (Aisa Mustapha 2012). These questions consisted of questions extracted from Encarta log provided by Microsoft and Excite query log (Mohideen 2007) and were purposely designed to be semantically identical but syntactically different. List questions and context questions were introduced in TREC-2001 QA track. The participating system was expected to identify more than one answer for a question as the complete answer consisted of more than one element. Also these answers related to each other, hence belonged to the same context. The expected answer length was also reduced

to 50 bytes. The TREC website notes that these questions were taken from MSNSearch logs and AskJeeves logs. TREC-2002 QA participants were required to give a single answer for each question and answer ranking was used (Mohammed 2005). Factoid and definition questions were introduced in the TREC-2003 QA track. These questions were taken from the logs of Microsoft and AOL. TREC-2004 QA focused on more complex questions and attempted to answer questions like 'How' and 'Why' returning only a phrase not a document.

General research has been going on in this area for a long time but Text REtrieval Conference (TREC) competitions have led to more specific research being carried out for the past decade or so. TREC has had a track dedicated to QA since 1999 (Anon, 2005) and has received a great deal of attention from the Computational Linguistics and Information Retrieval research communities in the last few years (e.g. TREC 2001–2003, cited by Svenonius (1972)). The definition of the task, however, is generally restricted to answering factoid questions (Svenonius 1972) with a focus on open-domain systems.

In contrast QA systems aimed at handling questions within a closed-domain can benefit from understanding the *'true'* meaning of the word in the question stated using natural language (Schlaefer 2007). For instance the word 'Heavy Metals' can be a difficult term to understand in a general context. This makes more sense when observed in a closed-domain such as entertainment i.e. *Heavy Metals* is a musical group in the entertainment domain.

A new task - complex interactive QA (ciQA) was brought forward in the TREC-2006 QA. The objective of this task was to investigate the suitability of the participants to meet the complex information needs of interactive systems. TREC-2007 made use of the same dataset as provided in TREC-2006 but with addition that the questions could be asked about blogs as well as newswires. The participants were expected to be capable of handling poorly formed language – frequently in blogs, as compared to well-formed language as in newswires. TREC-2007 QA led to a new campaign called Text Analysis Conference (TAC).

The Cross Lingual Evaluation Forum (CLEF) was an equivalent campaign set up for European languages as an initiative for research and development in monolingual and cross-language information access domains. Since 1998 the NTCIR campaign has provided cross-language evaluation for Asian languages working on Information Access (IA) technologies including information retrieval, question answering, text summarization, extraction, etc. Technolangue-EQUER is a similar campaign dealing with QA in French.

Likewise the aim of the Question Answering track at INEX (Initiative for the Evaluation of XML retrieval) is to investigate how the technology for accessing semi-structured data can

be used for NLP tasks. QA@INEX mainly deals with factoid and definition questions over resources like Wikipedia (in English), and a passage is expected as the extracted answer. Moreover there is research that focuses on *Paragraph Retrieval* and present paragraphs as the answer to the question (Rodrigo, Perez-Iglesias et al. 2010, Xu and Klakow 2010). Also REsPubliQA 2010 proposed two tasks related to QA for evaluating QA Systems (Peñas, Forner et al. 2010). One of these tasks was to retrieve a *Paragraph* as an A*nswer*. The main focus of the QA@INEX track is to combine different types of questions in a single task (Aisa Mustapha 2012).

Kuhn (1962) has persuasively argued that science progresses by means of its models of the general nature of a research area (cited by Shah 2010) — hence the motivation of TREC (QA) was also to design systems that can potentially produce fact-based short-answers (50 bytes long) to more generic questions. It also reflects the information access paradigm shifting from keyword-based search to natural language driven navigation (Mufti 2013).

General QA Architecture

A typical QA system performs question analysis, document/passage retrieval and answer selection in a sequential manner. These together constitute a single processing stream (e.g. Saad, Salim et al. 2011). This general pipeline architecture can be shown as in the block diagram shown in figure 2.4 (produced by myself):

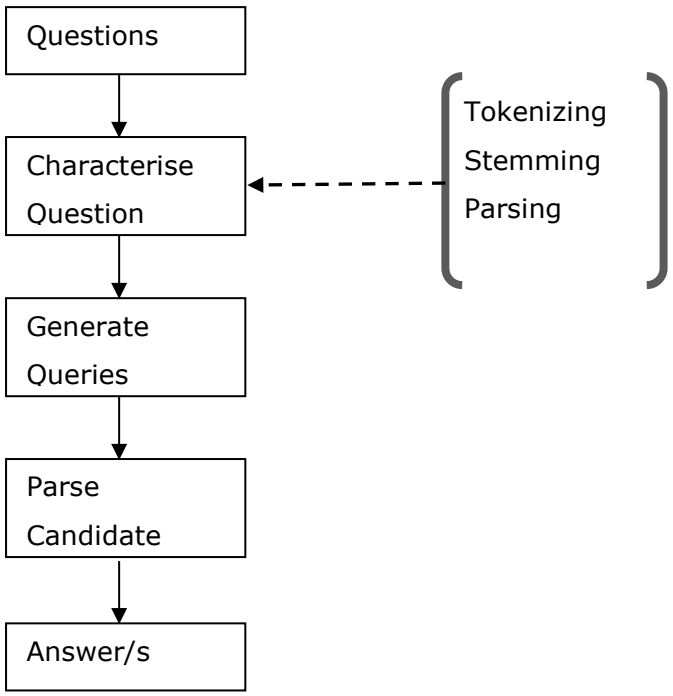


Figure 2.4: Components of a typical Question Answering System (Jilani 2012)

Some QA systems perform deep parsing while characterising a question. Others make use of ontologies to understand the context of the question better. Machine Learning techniques are used to learn the existing question patterns – pattern matching tables and named entity recognition tools are used for question annotation. Scholbash et al. (Meenachi and Sai Baba 2012) informs us that many open-domain QA systems start by harvesting a huge number of candidate answers by generating queries. Some approaches reformulate these queries in an attempt to target more information.

Another method is Question typology that involves determining *expected* answer type and/or deciding whether the candidate answer is of the semantic type expected by the question. In such approaches information combination is performed at this stage. Valero et al (Jilani 2013) indicate that common methods used for answer selection are based on lexical overlaps and answer redundancies (e.g. Anon 2002) and methods based on knowledge intensive methods (e.g. Mohd Yunus 2008). Finally parse trees of the question and candidate answer are compared to extract the potentially final answer. Another experimental approach used performs a form of reverse engineering. This is done by formulating potential questions for the free-text. This part of the process is done offline and stored for later use. These questions are later compared with the user query to provide an answer.

A block representation of components used by a typical Question Answering system has been presented in figure 2.5. This figure is based on the personal understanding of the author in order to provide nutshell view of a general QA system. This model has also been used later in this section when studying existing QA systems to aid quick understanding.

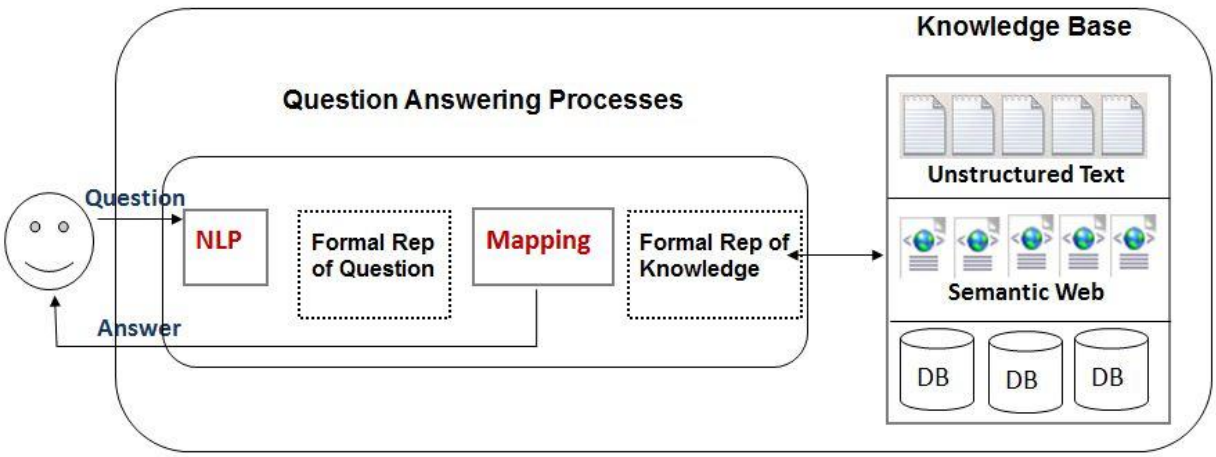


Figure 2.5: Block Diagram of a Typical Question Answering System (Jilani 2012)

All components presented in this figure may or may not be named and separated explicitly within each implementation of the QA System. The above block diagram has been reused to describe existing variants of the QA system. A brief outline of the sub-components presented in figure 2.5 has been included to enable further understanding:

NLP Processor – An NLP Processor is used to parse an input NL question/sentence. It can be further seen to have sub-components like:

- Sub-component to *tokenize* the sentences into smaller logical sentences
- Sub-component to *tag* the parts of speech for each sentence
- Sub-component to identify occurrence of *known* objects in the domain
- Sub-component to understand the *morphology* of the words

Multiple tools and techniques are available to implement this component but, again, the choice depends on the nature of the project under development.

Formal Rep. of Question – Based on the output of the NLP Processor user question, this is *represented* formally in machine-understandable format.

Formal Rep. of Knowledge – This component produces the *formal representation* of the candidate answer(s) for the question.

Mapping – This component is responsible to *map* a user question to an answer. Mapping is done by applying different comparison techniques. A successful mapping produces answer(s) to the question.

Knowledge Base – The knowledge base is the source of knowledge likely to contain the answer to the question. This knowledge can be structured in the form of a corpus such as the British National Corpus (BNC). It may be in the form of unstructured text such as on the World Wide Web or it can be constructed in machine understandable forms such as Databases and Semantic Web.

Following figure 2.6 presents architecture of a typical QA System suggested by (Keller 2013):

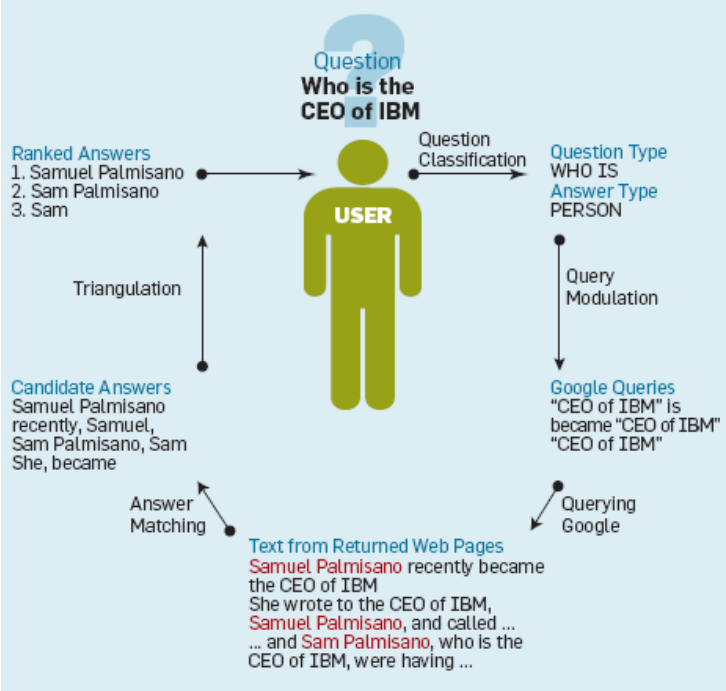


Figure 2.6: General Architecture of a Question Answering System (Keller 2013)

Recent Question-Answering Systems

The QA goal (of a typical NLP system) is to locate, extract, and provide specific answers to user questions expressed in natural language (Keller 2013). The following is a brief introduction to a few systems designed in an attempt to achieve this identified goal.

I – QA Model Applied to Document Retrieval System: 2009

A recent QA system prototypes the model proposed by Dang et al. (Kaplan 2001). It allows users to search the library by using simple natural language queries. These questions relate to the book's title, author, publisher and/or other (known) attributes of the eBooks. A series of steps (parsing, POS tagging, mapping) are performed on the user's query to understand it and then generate an appropriate answer. A pre-processing step trains the system providing a large number of user queries that are analysed to produce a vocabulary about the eBooks information. This vocabulary is categorised according to different POS tags and used to produce different syntactic structures (based on syntax of the query e.g. noun + verb + noun) and the semantic structures. Figure 2.7 presents an overview of the system architecture of Dang et al.'s proposed system.

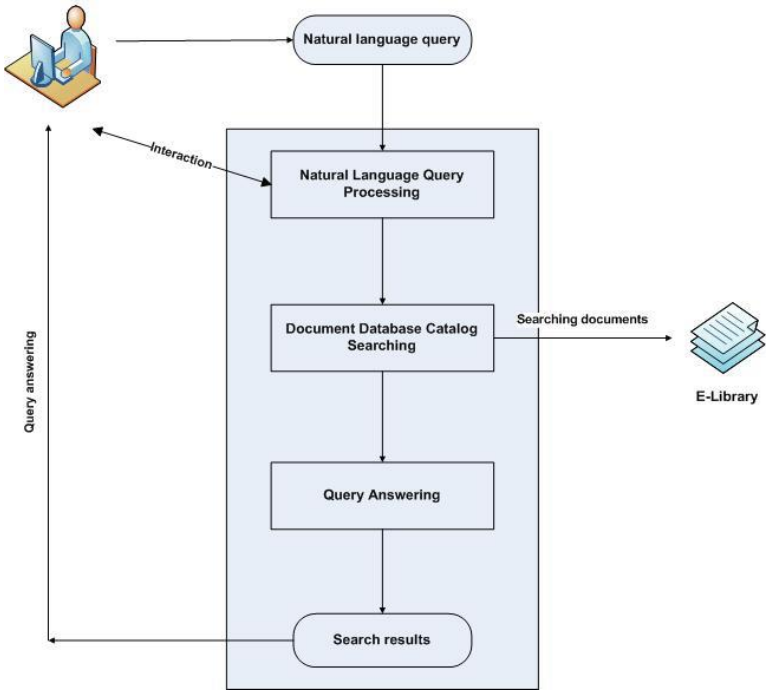


Figure 2.7: Overview of the System Architecture (Kaplan 2001)

Results of the pre-processing steps are used by the core component of the system to understand the user query. The user query is parsed, tagged and matched to the learned vocabulary. Unmatched tokens are ignored. Once the words are recognised, they are mapped to syntactic structures and transformed into semantic structures. Once semantic structure is determined, the rest of the process appears to be relatively trivial and deterministic. It involves generation of SQL queries to be executed against a database. The Database query result is then used to prepare the answer.

II – AquaLog: 2005

AquaLog is an ontology-portable QA system for semantic web. It consists of a linguistic component that is used to parse and classify a natural language input query and represent it as a Query-Triple <subject, predicate, object>. These query-triples are generated by using GATE annotations for a given input query. AquaLog also associates additional features to the query triple using JAPE grammar. Figure 2.8 provides a nutshell view of the AquaLog.

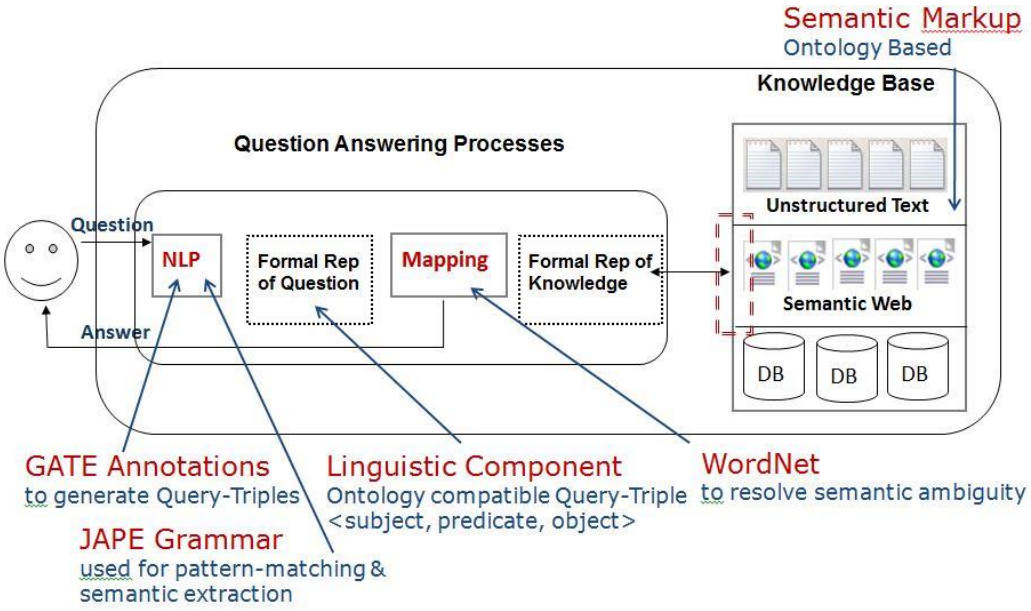


Figure 2.8: Overview of the AquaLog QA System (Jilani 2012)

The result of the linguistic component is translated into ontology-compatible triples using relation similarity. These ontology-compatible triples are ultimately mapped to semantic mark-up. Different mechanisms are applied to resolve any ambiguity including using WordNet or, as a last resort, asking the user to select the appropriate answer among different choices. User interaction could be a good option here but may not be suitable in some QA systems due to the nature of communication, e.g. mobile text messaging.

III – The NSIR System: (2000 - 2005)

NSIR (pronounced "Answer") is a web-based question answering research prototype system designed by the CLAIR group at the University of Michigan. It is one of early open domain QA system developed by an academic institute and it's underlying algorithms have been published in multiple publications. NSIR uses existing web-based search engines e.g. Yahoo, Google to retrieve the documents which are likely to contain the answer. It extracts entities by shallow parsing and returns a small taxonomy of answer types. NSIR uses trainable classifiers and Brill's POS tagger to identify certain words in the sentences that help determine the semantics of the question which in return helps in determining the semantics of the expected answer. Before answers are returned to NSIR users, they are ranked according to a set of machine-learning techniques, including the proximity algorithm and probabilistic phrase ranking (Ammon 2001) cited by (Jones 2004). The proximity algorithm ranks the answer *closer* if the words in the answer appear to be *close* to the words in the actual question. Probabilistic phrase ranking considers the part-of-speech tag sequences of the answer while ranking the potential answer.

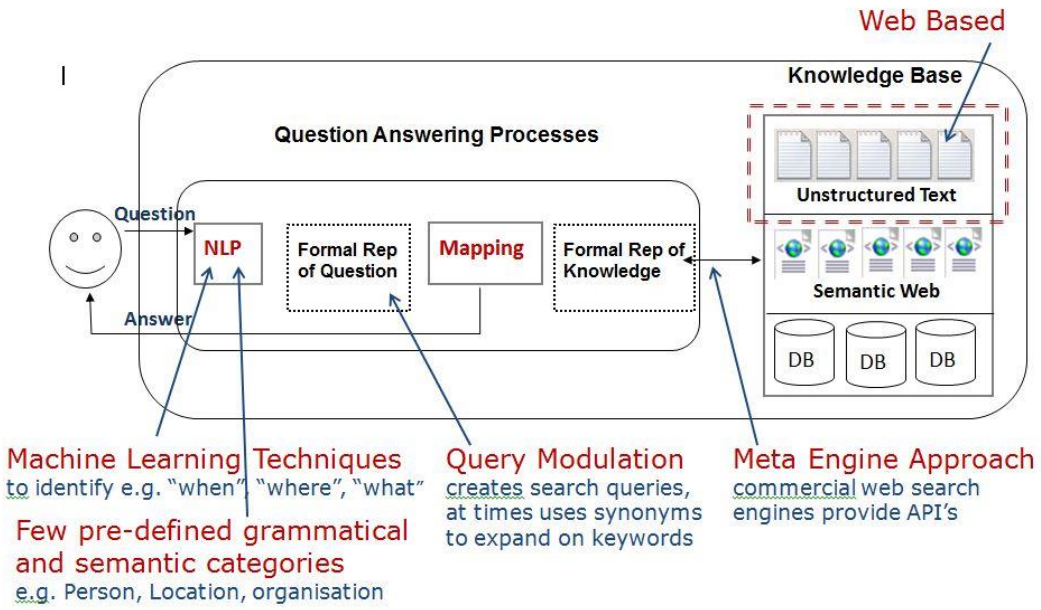


Figure 2.9: Overview of the NSIR Question Answering System (Jilani 2012)

Research experiments have shown NSIR capable of returning exact answers or snippets of text containing the answer (Keller 2013). The above figure shows a nutshell view of the NSIR QA System.

IV – ASU QA: 2002

An article in an online magazine *Information Week*, dated 28th March 2005, mentioned the ASU QA system as one of the most promising directions in the "Search of Tomorrow" (Mufti 2013). ASU QA is a research prototype designed by Arizona State University. It uses trained regular expressions to identify certain grammatical categories (e.g. "what is", "who is") of the user question. Instead of using typical tools ASU QA uses a pattern-matching mechanism for named-entity-identification in extracted answers. Once the grammatical type of the question is identified, ASU QA exploits the "meta-engine" approach to retrieve potential answers by sending multiple queries to existing search API's. These queries are produced by using dozens of pre-trained question patterns it has learnt through training while answering similar questions in the past. ASU QA uses the probabilistic triangulation technique to avoid redundant answers retrieved from the web.

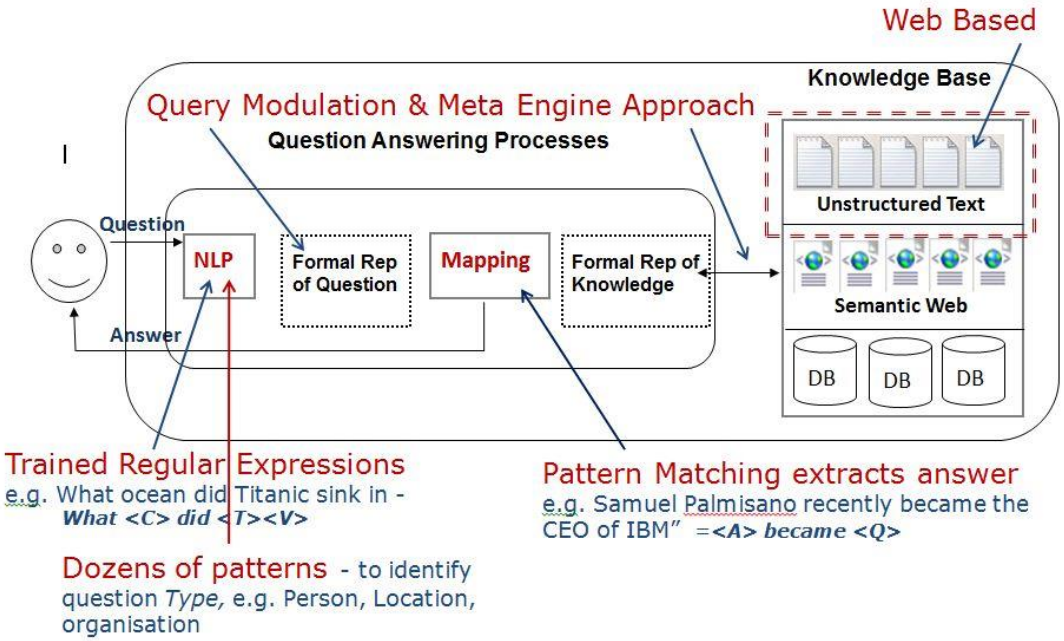


Figure 2.10: Overview of the ASU-QA System (Jilani 2012)

Figure 2.10 gives a nutshell view of the ASU-QA System. ASU QA was demonstrated in 2005 to help locate potentially malevolent online content, potentially helping law-enforcement (Keller 2013). Alongside other research tools ASU QA has been used for several years for research purposes in a $2 million project supported by NASA.

V – Falcon: TREC9 (2000)

Falcon has been identified as one of the most successful systems presented in the TREC9 (2000) track (Mufti 2013)that makes use of open domain linguistic resource WordNet. NSIR and ASU QA only use grammatical and semantic *types* whereas Falcon goes an extra step; it classifies the identified words in the question in pre-built semantic *types* (like person, place, profession, date, etc.) and *sub-types* (such as musician, writer, politician etc.) in an attempt to understand the semantics of the user question which in return helps in identifying the potential *correct* answer to the question.

Falcon uses purpose built Information Retrieval (IR) techniques to find useful paragraphs of text within the documents. Figure 2.11 gives a nutshell view of the Falcon QA System. Falcon uses automated surface-based NLP methods such as the loose unification technique for syntactic parsing instead of the traditional bag-of-words approach to validate the expected answer and to eliminate incorrect answers. Instead of operating at word level, Falcon operates at the level of *dependencies between words* which provides better extraction of question-and-answer semantics. Justification for the correct answers is achieved by translating the semantic forms into logical forms, relying on minimal abduct knowledge (Morato, Marzal et al. 2004).

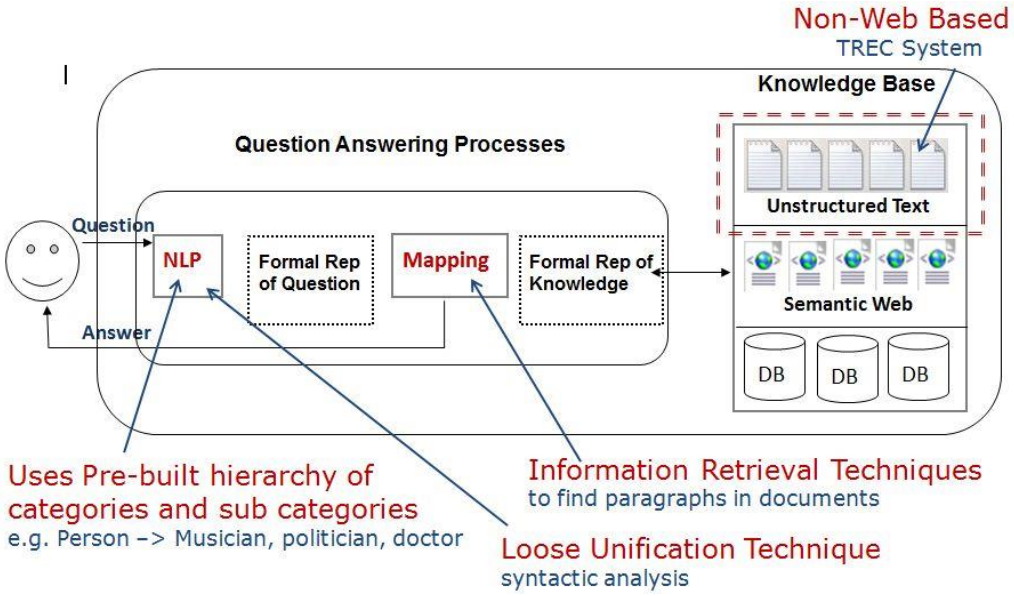


Figure 2.11: Overview of the Falcon Question Answering System (Jilani 2012)

VI – QA-LaSIE: TREC9 (2000)

QA LaSIE is an IE system designed for research purpose by Sheffield University. It is an extension of the LaSIE (Large Scale Information Extraction) system that was originally designed for participating in message understanding conferences (MUC6 & MUC7). As LaSIE did not have grammar rules to handle sentence-posing questions this feature was added on; hence the name QA LaSIE. This capability was achieved by introducing a special semantic predicate module to identify entity under question.

The user question is passed to the Information Retrieval (IR) system. IR compares this question against a repository of (unstructured) text documents to find candidate passages. IR takes a number of documents to return and the number of paragraphs in the passage as parameters. The question with the passages identified by IR is sent to the QA-LaSIE. These are parsed by LaSIE, which internally uses GATE for tokenizing, tagging, etc.

The candidate answers returned are ranked by assigning them scores which are filtered on multiple assumptions such as "in most cases it is unlikely that a correct exact answer to a question will contain many, if any, of the non-stopwords in the question"(Abbott and Ryan 2000), before the final answer is returned. The NLP research group at Sheffield University has a special interest in understanding the role of IS systems and they encourage retrieving a small set of candidate answer-bearing documents to be intensively analyzed by a second stage answer extraction component (Anon, 2009).

VII – AskMSR

AskMSR is still a prototype, although Microsoft is trying to improve it and it may be launched commercially under the name AnswerBot (Twilightdreamlover 2008). AskMSR uses the web as a data repository and depends on data redundancy for generating answers to questions. The AskMSR team stresses achieving harder things by simpler means. They claim that it is more difficult to find the *correct* answer for a question where the answer is present in a small data set. They are of the opinion that in a small data set, e.g. the TREC corpus, finding the *correct* answer would require the application of complex syntactic and semantic techniques. On the other hand the Web is a huge data repository and the *correct* answer might occur redundantly in several simpler forms. If redundancy is worked around by applying probabilistic algorithms, it can assist in finding the correct answer, hence making the complex task easily achievable. Figure 2.12 gives a nutshell view of the AskMSR System.

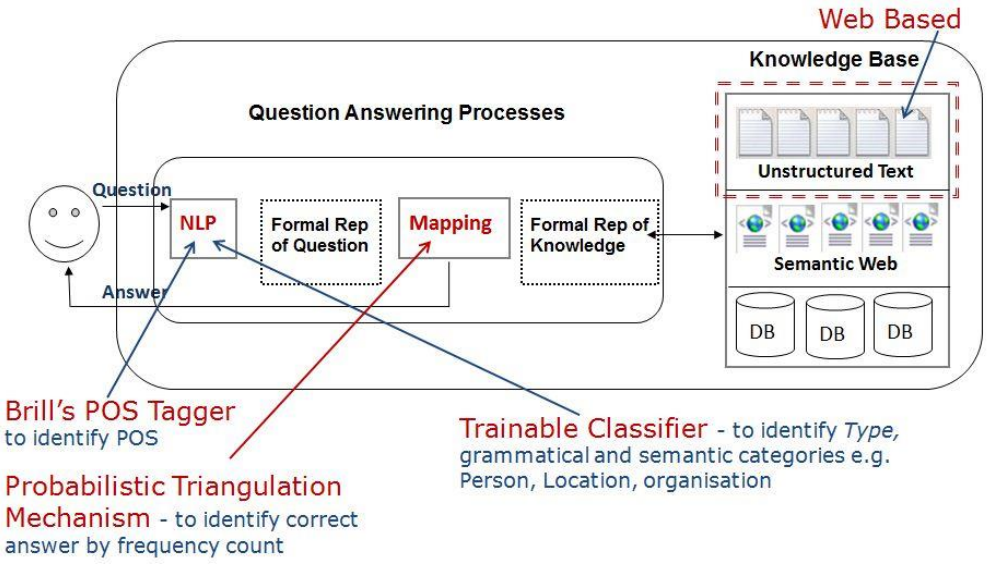


Figure 2.12: Overview of the AskMSR Question Answering System (Jilani 2012)

AskMSR generates several rewrite strings for the question assuming them to be sub-strings (sometimes by just ANDing the words appearing in the questions) of the expected *correct* answer. Question-type determination has been widely used (Brill et al., 2001). Question transformation and question expansion are also commonly used, via manually crafted question-to-query transformations (Brill et al., 2001). AskMSR does not use a parser or POS tagging for query formulation but uses a lexicon at times (Jones 2004) to identify the morphological equivalents for the words in the substrings. These question-string rewrites are sent to the search engine and returned document summaries are searched through in hope of finding a matching pattern. AskMSR does not search through the complete document for efficiency reasons. Because a wrong answer is often worse than no answer

(Jones 2004), AskMSR attempts to identify situations where the system is more likely to return an incorrect answer.

Eric Brill (Yang, Newby et al. 2005) claims that AskMSR is a novel system as state-of-the-art accuracy is achieved using very simple methods whose power comes entirely from the plethora of text currently available to these systems, as opposed to deep linguistic analysis or the application of state-of-the-art machine learning techniques. He also suggests that the field of NLP might benefit by concentrating less on technology development and more on data acquisition.

VIII – START: 1993

START (SynTactic Analysis using Reversible Transformations) was developed by Boris Katz at MIT's Artificial Intelligence Laboratory (Anon, no date). The START official website states that START works by parsing the incoming questions, matching the queries created from the parse trees against its knowledge base and presenting the appropriate information segments to the user. The START team claims that "this way, START provides untrained users with speedy access to knowledge that in many cases would take an expert some time to find" (Anon, no date). START in contrast to other QA systems has the capability to bring up images whose annotations can be matched to the question.

START learns from analysing information presented in English. This information is then stored in the Knowledge Base as a *digested summary* of the syntactic structure of the sentences learnt. The question sentence is analysed in a similar way and the results are matched with the information stored in the knowledge base. This matching is performed using sophisticated linguistic techniques and syntactic patterns, hence achieving more than just keyword matching. START, if not able to find an answer, does not offer any web links with potential answers as some systems would do. START fails in differentiating between sentences which have closely related meanings but different surface structure. The START team is working to enable the system to handle more complex questions like – "Who is the president of France" (Padesky 1993).

Multi-Stream Question Answering

The information revolution, due to the widespread of the Internet, has brought several new areas of research into focus. Automatic Text Processing is a vital area in this context. Hence automated Question answering has become a "wanted" functionality that current search engines like Google and Yahoo are "aiming" to achieve. START and DFKI are the experimental web-based QA systems that have pioneered this attempt. Regardless of the present and past efforts made in the area, the use of QA systems on the web is not very

visible as compared to common *keyword* search engines at the moment. One of the contributing reasons is the fact that current QA systems do not have the same performance level across different languages. The best performing system in TREC 2004 successfully answered 77% of factoid questions in English (Hartman 2008) whereas a similar system meant to answer similar kinds of factoid questions in Spanish only achieved 55% accuracy (Magnini et al, 2006).

This variation of performance across languages has led to research into a more recent approach; what is termed Multi-Stream Question Answering. Valero (Aisa Mustapha 2012) explains that "in this approach the idea is to combine different QA strategies in order to increase the number of correctly answered questions".

Multi Stream QA architecture has also been advocated to address other challenges faced in language processing tasks like recall problems occurring due to vocabulary gaps, "the phenomenon that the question and its answer(s) may be phrased in different vocabularies" *(Xu and Klakow 2010)*. Multi-Stream QA can be either Internal or External. It is Internal if the multiple strategies are applied to the components of a QA system as in the work of Pizzato (2005) and Chu-Carroll et al (2003). The approach is External if it makes use of the finalised output of multiple QA systems and employs adapted strategies of well-known information fusion techniques to enhance the answer quality, based on the ideas proposed in the earlier works of Diamond (1996) and Vogt et al. (1999) in this respect, Valero et al (Aisa Mustapha 2012) have classified these adaptations of the *External Multi Stream QA approach* into five categories:

I – Skimming Approach

In this adaptation the answer streams of different QA systems are skimmed through based on the ranks set by individual QA systems. The works of Clarke et al (2002) and Jijkoun et al. (2004) involve use of this approach. Two variants of the approach are proposed:

(i) Naïve Skimming Approach: The answers from the multiple streams are selected without any preference or order i.e. at random.

(ii) Ordered Skimming Approach: In this approach answers are selected in preference of their overall *general confidence value* from the multiple-streams.

II – Chorus Approach

In this adaptation the answers are ordered based on their frequency of occurrence in the answer streams of different QA systems. In simple words, this approach uses redundancy to order the QA systems. Some systems using this approach have been referred to in the

works of Burger et al. (2002), Jijkoun et al. (2004), Raussivnov et al. (2005) and Rotaru et al. (2005).

III – Dark-Horse Approach

This approach orders the answers from multiple streams according to their general confidence value for each type of question. In simple words, this approach produces separate confidence value given to factoid, definition or list type of questions. Jijkoun et al. (2004) refers to a system based on this approach.

IV – Web Chorus Approach

This approach ranks the answers based on the number of web pages containing the answer terms. This approach was introduced by Magnini et al in 2001 and was further evaluated by Jijkoun et al. (2004).

V – Answer Validation Approach

As described by Valero et al. in this approach the correct answer is decided based on its entailment with a given support text. This was proposed by Penas et al. (2007) and implemented by Glockner et al (2007).

Hybrids of the above approaches have also been in use and have been reported to produce improved results e.g. Jijkoun et al. (2004). Valero et al. have made use of the output of multiple QA systems that had weak performance as individual systems. Evidence has shown that using this approach potentially adds up to 31.5% accuracy in answers (Valero et al. 2010). The proposed approach is claimed to use an answer validation method that confirms the answer based on the supporting text containing the answer. This approach does not rely only on the ranking by occurrence or stream confidence-level alone and has yielded a 19% relative increase in accuracy as compared to the best participating QA system in TREC.

Dalmas and Webber (2006) agree that in the context of QA systems, "Much has been written about the qualities of a good question but little about the qualities of a good answer" (Ely, 2002). Mostly QA systems are expected to bring up only one or *N* right answer(s) whereas this is not the case always. In some cases there is more than one right answer for a given question. As the research in automated QA stands, it has led to the impression of treating candidate answers as competitors to each other and not as potential allies (Dalmas, Webber 2006). It is believed that the relationship between answers can be exploited to produce higher quality answers. Also it is important to note that the information available across multiple answers can be fused together to present a single comprehensive answer.

Another aspect of research is that it requires materials existing in multiple documents to be brought together at a point indicating where they exist in them. Of course this also needs a set evaluation mechanism in place that defines what is considered as relevant information to compose an answer and what is loose information. This is not to be confused with the documents being returned by performing keyword search, as in popular search engines. This situation is more evident when dealing with list questions as in TREC-2001 but it is also true for factoid questions that can have more than one correct answer. Consider the factoid example question where more than one answer presenting different information is returned and fusion of this information can return a more comprehensive and useful answer for the question below:

Q: Who is Qaid-e-Azam?
A1: Qaid-e-Azam is the title given to Muhammad Ali Jinnah.
A2: Qaid-e-Azam is the founder of Pakistan.
A3: Qaid-e-Azam was the first president of Pakistan.

Recent research indicates the value of semantic relations between candidate answers for answer selection and emphasises that these answers should be seen as allies to each other (Anon, 2011). Research indicates that a QA system's performance can potentially be improved by:

- Taking advantage of multi-stream QA systems approach benefits
- Offering a more *useful* and *rich* answer to question

Multi-Stream QA systems use filtering mechanisms to remove the wrong answers from the candidate answers. Based on all of these potential answers being correct answers, Moriceau (2005) classified these relationships under four categories as follows:

Equivalence: This exists where the answers are lexicographically different answers but they represent the same entity, e.g. notational variations, synonyms or paraphrased answers.

Q: When does the ghost appear in Shakespeare's drama Hamlet?
A1: midnight **A2:** 12am

Inclusion: This exists where more than one answer are consistent but differ in the detail of information they present, e.g.

Q: What role did Hamlet have in Shakespeare's drama Hamlet?
A1: The prince **A2:** The heir **A3:** The oldest son

Aggregation: This exists where the answers are consistent but if the information they present is fused they can provide a more entailing answer in conjunction, e.g.

Q: Who was Hamlet in Shakespeare's drama Hamlet?
A1: Future king and son of Gertrude **A2:** Accused king's nephew and late king's son

Alternative: This exists where the answers are not consistent and normally provide unique information that cannot be presented in conjunction with each other, e.g.

Q: Was Hamlet mad in Shakespeare's drama Hamlet?
A1: Hamlet was mad **A2:** Hamlet was pretending to be mad

The experimental findings of recent research show that making use of simple techniques to recognise these relations can boost the accuracy of the answer by 10% (Anon, 2011).

Critical Analysis of QA Systems

Early statistics show that, "Question Answering (QA) systems can answer nearly 85% of factoid questions" (Di Cai and Lee McCluskey 2011). Yet none of the above reviewed systems benchmarks a QA system probably because each of them attempts to explore and experiment different aspects of QA. Dang et al (Kaplan 2001) comment that out of the QA systems, such as START, AskMSR, and NSIR, START may be considered as the best system in returning good answers for users. However, START is only able to answer questions about concepts, but it cannot answer questions about causes and methods (Kaplan 2001). Figure 2.13 gives a nutshell view of common techniques and methods QA systems use for question answering in general as identified by the author during this research work.

The approach of using *unstructured text* within existing documents (e.g. QA-LaSIE) has the potential to benefit from available resources of knowledge that have not been specifically designed for this purpose. This approach works where the nature of the knowledge source is *static* (e.g. text documents or static web pages). But it cannot be applied to *dynamic* knowledge sources (e.g. dynamic web pages) that are place-holders only and actual information is stored in backend data-stores and made available only at run-time. QA systems designed on this approach can potentially answer any generic question because of their open-domain nature. However, as these systems have to process large volumes of unstructured text, researchers argue that "deep text analysis becomes a resource-consuming task" (Schlaefer 2007).

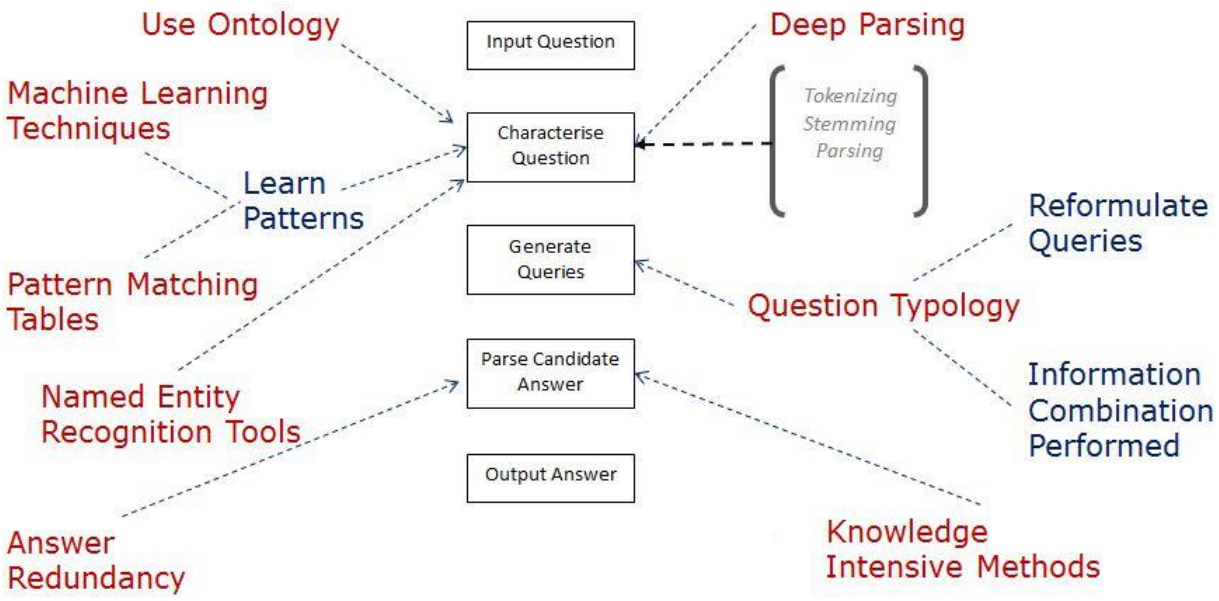


Figure 2.13: Methods and Techniques used to achieve effective Question Answering (Jilani 2012)

A study (Keller 2013), conducted to compare few popular web keyword-based search engines with known QA systems (including NSIR, ASU QA, Falcon), concludes that open-domain QA has emerged as a technology that complements or even rivals keyword-based search engines. However, to compete with established keyword-based search engines, QA systems still must address several technical challenges such as *scalability, credibility* and *usability* (Roussinov et al, 2008).

Section Summary

Question-Answering (QA) is an important research area, which is concerned with building a system that automatically answers questions posed by humans in a natural language. A technical review of different QA system models and methodologies reveals that a typical QA system consists of different components to accept a natural language question from a user and deliver its answer(s) back to the user. These components are used to *understand* the user question by passing it through successive NLP processes and then *represent* it in machine-understandable format. This representation needs to be compatible with the knowledge representation so that it can be mapped to an answer using different mapping techniques available. This mapping can be probabilistic, deterministic or a combination of both QA systems and is built on top of a knowledge base that is used to extract the potential answer(s) to the user question. This knowledge base is either built specifically for the system (domain-specific knowledge base for close domain systems) or an existing resource (such as the World Wide Web or a text repository) can serve as a knowledge base (domain-independent knowledge base for open domain systems). Different tools and

techniques are available to implement each component but their selection is influenced by the nature of the system under design.

Despite all the up-to-date research in the subject area, researchers still agree that "parsing a sentence written in natural language is as yet one of the most difficult computing challenges and understanding the meaning of the sentence is proving even more difficult" (Sanderson and Alberair 2001). Having said this, it is also true that "several well-known futurists believe that computers will reach capabilities comparable to human reasoning and understanding of languages by the year 2020" (Roussinov et al, 2006).

A notable fact for the above discussed QA systems is that none of these has been tested on religious text with the aim of question answering. These systems have been aimed at structured/ unstructured data collected from everyday English text i.e. text collected from television programs, news wires, conversations, novels and other like genres. The process of QA on religious data can be considered as finding a subsection of the religious corpus that satisfies the need of the query. This thesis attempts to apply a combination of advanced Information Retrieval and Natural Language Processing tools and techniques to develop a Question-Answering system for religious QA. The Quran, the core text for the religion of Islam, has been chosen as the first case study. The next chapter describes the Quran as a document and corpus to enable readers to understand the peculiar nature of this document and the challenges associated with Question–Answering on this corpus.

2.2 The Quran Corpus

Recently the Quran has been considered as a potential subject for computing research in corpus linguistics (Dukes 2011), Text Analysis (Abdul-Baquee 2013) and natural language processing (Eric Atwell 2013). Research in these subject areas is cross disciplinary as it involves knowledge of several domains and applications. Eric Atwell informs us "in general Natural Language Processing research involves Machine Learning from a domain-specific corpus of text documents". The following are a few attributes that make the source text corpus ideal for computational analysis:

- The text corpus should have no privacy constraints and costs involved. It should be available free for research purposes.
- A large community of experts for that corpus exists, which has developed 'linguistic tagging schemes' or ontologies for the domain.
- A large group of users exist, that can assist in evaluating the system and its results.
- A huge group of potential users exist, so research in the domain has an impact

The Classical Arabic Quran meets the above listed criteria in all respects as there are over a thousand years of documented work on the linguistics of the Classical Arabic text of the Quran. Besides most features and resources mentioned above are valid for the English translations of the Quran also. In today's global population Muslim readers of the Quran comprise "the largest user-group ever for a single text corpus" (Atwell 2013). Out of this huge population, the audience user group for research on English translations of the Quran is still a huge group, i.e. the billions of Muslims worldwide. Quranic Arabic corpus linguistic research has produced ontological resources and tools such as the named entity ontology (Dukes 2011), semantic concept "tags" (Abbas 2009), QurAna (Sharaf and Atwell 2012) and QurSim (Sharaf and Atwell 2012). Although these resources were developed for the Arabic Quran they remain equally useful for research on English translations of the Quran as the verse references are the same regardless of the language of the Quran.

This chapter is divided as follows. The first section provides background for the Quran as a document as well as a corpus. This section provides an insight into the historical information, the structure, the language and prominent topics of interest for the user of the Quran. The next section details the peculiar characteristics of the Quran. A brief introduction to the English translations of the Quran is given in the following section. A discussion about why there is a need for Question-Answering research on the Quran and what impact and value it could have for Muslims is presented. Lastly, some research work is received that involves the use of corpora derived from the religious books of Christianity, Hinduism and Sikhism.

I – The Quran

This section provides some background knowledge about the Quran. This includes the history of the Quran, the structure of the Quran and the language used in the Quran.

(i) History of the Quran

The original version of the Quran is in Classical Arabic. It is a text of less than 80,000 Arabic word (Atwell 2010). The followers of Islam, Muslims, hold that the Quran is composed of the direct words of God (Allah) revealed gradually to the last prophet of Islam, the Prophet Muhammad (S) by Angel Jibrail. The Prophet Muhammad is also known as *The Messenger* of Islam. The Quranic verse states:

> *"It is He (Allah) who has sent among the illiterate a Messenger (Prophet Muhammad) from among them, to recite His songs to them, and to purify them and to teach them the Book and the Wisdom." 062:002*

Researchers studying the Quran look at the claim of revelation closely connected to the name given to the book itself. Cynthia Whissell (Whissell 2004) studying the emotion and styles of language in an English translation of the Quran notes that the word 'Quran' can be translated literally as 'recitation'. She indicates that it is not surprising that the transmission of the Quran was predominantly oral rather than written. She proves this by the fact that most 'revelations' instructed the Prophet Muhammad to "Say!" or "Recite!".

> ***"Say** to (them): "If the home of the Hereafter with Allah is indeed for you specially and not for others, of mankind, then long for death if you are truthful."*
>
> *002:094* *Translated by Muhammad Muhsin Khan*

> *"**Say**: Allah knows best how long they remained; to Him are (known) the unseen things of the heavens and the earth; how clear His sight and how clear His hearing! There is none to be a guardian for them besides Him, and He does not make any one His associate in His Judgment."*
>
> *018:026* *Translated by MH Shakir*

> *"And **recite** what has been revealed to you (O Muhammad ﷺ) of the Book (the Qur'an) of your Lord (i.e. recite it, understand and follow its teachings and act on its orders and preach it to men). None can change His Words, and none will you find as a refuge other than Him."*
>
> *018:027* *The Noble Quran*

The Prophet Muhammad recited the 'revelations' to the initial followers of Islam (Companions), who memorised them. This learning remained oral for a long time until finally it was transcribed verbatim by the early Muslims. The first canonical written version of the Quran was authored by scholars in the Uthmanic recension on the order of the third Caliph of Islam. This compilation has remained the same since then. Nevertheless there still are seven lawful reading versions of the Quran. However, the differences are minor and residual in nature so do not provide enough grounds to believe that Quran has 'versions' in the strict sense of the word. The differences between these legitimate compilations are mainly in the manner of oral recitation and interplay between the recited and written form of the Quran (Baker 2008). The Quran is believed to be a single-authored text and accepted by Muslims as being the (primary) authoritative source of knowledge, wisdom and law (Atwell, Brierley et al. 2011).

(ii) Structure of the Quran

The Quran is compiled as 114 chapters (Surahs) of unequal length. Each chapter is named based on different criteria. The order of this compilation is widely acknowledged to be different to the order of revelation which has a direct relation with the calendar of Islam. Chapters are classified based on the time and place of revelation.

- Maccan – revealed during the stay of the Prophet Muhammad in Macca
- Madinan – revealed during the stay of the Prophet Muhammad in Madina

Generally Maccan chapters precede those revealed in Madina, which appear later in the compilation of the Quran. There are different viewpoints relating the order of chapters in the Quran. The contents of the early Maccan chapters are more 'poetic' in nature than the later Madinian ones. Chronologically, the order of chapters describes the Prophet Muhammad as a public warner, religious historian, guardian of the faith and finally in Madina, as a warrior and legislator (Whissell 2004). According to another theory the chapters presenting law and structure precede those that relate to awe and insight. The shorter chapters tend to appear towards the end of the compilation as compared to the longer ones which make up the earlier part of Quran. But an exception to this is the first chapter Al-Fatiha which is shorter than the following chapters. Whissel notes that "current order of chapters, as well as the specific order of materials within them is attributed to the Prophet Muhammad himself, and it is considered to be meaningful" (Whissell 2004).

The chapters of the Quran have been grouped together making multiple logical sections to facilitate the reciting purposes of the Muslims. The smallest unit of the Quran is the Verse (Ayah). Scholars have different opinions on the total number of verses in the Quran. This number ranges between 6,214 and 6,666 verses in total but the content of the Quran remains the same across all these calculated figures. This difference of opinion is due to the different punctuation of longer verses. Some scholars break these into more than one smaller verse for their calculations. Verses are of unequal length and their number varies amongst chapters, e.g. the chapter Al-Kausar is the smallest being 3 verses in length and the chapter Al-Baqara is the longest with 286 verses, verse number 282 being the longest verse in the Quran. Scientific calculations indicate that the Quran corpus consists of 77,784 word tokens and 19,287 word types (Sawalha and Atwell 2010).

(iii) Language of the Quran

The Quran is written in Classical Arabic (CA). Most scholarly literature on the Quran authored between the 7th and 10th centuries has used Classical Arabic. Historically Classical Arabic has its roots in the early ages of Islam (Versteegh 1997) hence it has developed an inseparable cultural link with Islam. Chronologically Classical Arabic has remained functional for over fourteen centuries since the rise of Islam. Going through an evolution process Classical Arabic has developed into Modern Standard Arabic (MSA) to adjust to the linguistic needs of the modern age. Arabic is a language having complex morphological structure (Muhammad 2013); (Attia 2008). MSA also exhibits features of diglossia (Ferguson 2006). MSA is a strategically important and interesting (Bakalla 2002) language as it is spoken by millions of people (PEWResearch 2011). It is spoken by inhabitants of countries vital to the world economy because of their oil reserves. Also over 1.4 billion Muslims use the Arabic language to offer their prayers several times a day (Farghaly and Shaalan 2009).

Despite the development of Arabic as a language of current use, the language of the Quran has remained unchanged. Native speakers of Arabic regard the Quran as the linguistically and stylistically *unparalleled masterpiece* of Arabic.

(iv) Prominent Topics of Interest in the Quran

Muslim scholars claim Islam to be a *'Deen'* and not a mere *'Religion'* (Svenonius 1972, Di Cai and Lee McCluskey 2011)). The notion of 'Deen' goes beyond a narrow conception of religion to encompass notions of a civilised life and the intellectual heritage of a society. Islamic scholars insist that the Quran has rulings and knowledge to guide almost all subjects of life (Oktar 2013). A Muslim is expected to learn these rulings and implement them in order to become a responsible member of Muslim society. Susan et al argue that "In learning about the deen, they (Muslims) would learn to carry out the duties of the faith, and to act according to its principles" (Jones 2004).

The Quran is the core document providing fundamental knowledge of Islam. Muslims need to refer to it frequently to learn and practice Islam. No research was found which investigated the prominent topics of interest as seen by the Muslim community. To fill this gap, we have assumed the Islamic information dissemination services as a benchmark. Details on these websites are provided later in Chapter 4, section 4.4. The rationale behind this is that questions asked by Muslims using Q&A services on these Islamic websites give an insight into the practical aspects of Islam people want to know more about. Hence it can be argued that these are the topics of interest that need to be addressed. The author of this

thesis studied the patterns of these Q&A services and found several such areas. But only the following few have been taken up as pilot case study in this thesis. These can be classified as follows:

Rituals & Faith

According to Islam there are five basic 'pillars' of Islam; namely Shahada, Salat; Saum, Zakat and Hajj. The Quran provides core verses for each of these pillars. It is observed that scholars derive rulings around these pillars of Islam deduced from these core verses. The core verses for each of these pillars are given below.

(i) Shahada

The notion of "Shahada" or "Testament of faith", can be best expressed as "Belief in Allah". According to scholars, this concept has been derived from the following two verses:

> *"Allah bears witness that **La ilaha illa Huwa** (**none has the right to be worshipped but He**), and the angels, and those having knowledge (also give this witness); (He is always) maintaining His creation in Justice. La ilah illa Huwa (none has the right to be worshipped but He), **the All- Mighty, the All-Wise**."*
> *003:018* *Translated by Khan*

> *"When the Hypocrites come to thee, they say, `We bear witness that thou art the Messenger of ALLAH.' And ALLAH knows **that thou art indeed HIS Messenger**, but ALLAH also bears witness that the Hypocrites certainly are liars."*
> *063:001* *Translated by Sher Ali*

(ii) Salat

Salat is the Islamic form of ritual prayer. There are several verses in the Quran that refer to the importance and details of prayer in Islam. Ablution, which is very closely related to prayer, is cleaning or washing to get ready to perform prayer. The following Quranic verse gives the core steps involved in Ablution:

> *"O you who believe, when you observe the Contact Prayers (Salat), you shall: (1) wash your faces, (2) wash your arms to the elbows, (3) wipe your heads, and (4) wash your feet to the ankles. If you were unclean due to sexual orgasm, you shall bathe. If you are ill, or traveling, or had any digestive excretion (urinary, fecal, or gas), or had (sexual) contact with the women, and you cannot find water, you shall observe the dry ablution (Tayammum) by touching clean dry soil, then rubbing your faces and hands…."*
> *005:006* *Translated by Rashad*

(iii) Saum

The term Saum in Arabic refers to Fasting in Islam. This term has several equivalents and variants. These are discussed later in section 4.10. There are multiple verses in the Quran that provide details of different aspects of fasting. For example the following verse has been frequently referred to by scholars to emphasise the importance of fasting as a basic ritual in Islam.

> *"Believers', fasting has **been made mandatory for you** as it was made mandatory for the people before you, so that you may have fear of God."*
> *002:183* *Translated by Sarwar*

(iv) Zakat

Zakat is an Islamic form of taxation where rich Muslims are bound to contribute towards the betterment of less fortunate Muslims in the community. There are several grey areas in the calculation and payment of Zakat that Muslims consult scholars for. The concept has been referred in the Quran several times but the following verse emphasises that it is an obligation for a Muslim.

> *"Welfare funds (zakat) are only for the poor, the destitute, the tax collectors, those whose hearts are inclined (towards Islam), the slaves, those who cannot pay their debts, for the cause of God, and for those who have become needy on a journey. **Paying zakat is an obligation that God has decreed. God is All-knowing and All-wise.**"*
> *009:060* *Translated by Sarwar*

(v) Hajj

The holy pilgrimage in Islam is known as Hajj. This act of worship involves practical, logical and conceptual aspects. Muslims refer questions to scholars for guidance on these aspects. Along with other Islamic literature, scholars also reference the following Quranic verse in answering specific questions.

> *"And accomplish the pilgrimage (Hajj) and the visit for Allah, but if, you are prevented, (send) whatever offering is easy to obtain, and do not shave your heads until the offering reaches its destination; but whoever among you is sick or has an ailment of the head, he (should effect) a compensation by fasting or alms or sacrificing, then when you are secure, whoever profits by combining the visit with the pilgrimage (should take) what offering is easy to obtain; but he who cannot find (any offering) should fast for three days during the pilgrimage and for seven days when you return; these (make) ten (days) complete; this is for him whose family is not present in the Sacred Mosque, and be careful (of your duty) to Allah, and know that Allah is severe in requiting (evil)."*
> *002:196* *Translated by Shakir*

Natural Science

Besides other knowledge, the Quranic verses have been argued by some modern scholars to provide information on natural science. This includes several interesting scientific concepts such as Evolution of Earth, Doomsday, Human body, Concept of Paradise, Day & Night, principles of the Solar system, and Human Foetus. Some websites provide articles dedicated on the subject of Science in Islam (Oktar 2013).

The following few Quranic verses have been arguably observed being referenced by Islamic scholars in connection to the scientific knowledge available in the Quran.

> *"In the creation of the heavens and earth, and the alternation of the night and day, and the ships which sail the seas to people's benefit, and the water which Allah sends down from the sky — by which He brings the earth to life when it was dead and scatters about in it creatures of every kind — and the varying direction of the winds, and the clouds subservient between heaven and earth, there are signs for people who use their intellect."*
> *00 2:164*
>
> *"Surely the **number of months with Allah is twelve months** in Allah's ordinance since the day when He **created the heavens and the earth**, of these four being sacred; …"*
> *009:036* *Translated by Shakir*
>
> *"He **created seven universes in layers**. You do not see any imperfection in the creation by the Most Gracious. **Keep looking;** do you see any flaw?"*
> *067:003* *Translated by Rashad*
> *"The sun sets into a specific location, according to the design of the Almighty, the Omniscient. The moon we designed to appear in stages, until it becomes like an old curved sheath. The sun is never to catch up with the moon - the night and the day never deviate - each of them is floating in its own orbit. "*
> *036:038 – 040* *Translated by Rashad*
>
> *"He created the heavens and the earth in true (proportions): He makes the Night overlap the Day, and the Day overlap the Night: He has subjected the sun and the moon (to His law): Each one follows a course for a time appointed. … "*
> *039:005* *Translated by Yusuf Ali*

History

A part of the Quran presents historic information. This includes knowledge of pre-Islamic and post-Islamic events and incidents. The Quran mentions twenty-five Prophets across Quranic verses. This includes life stories of Prophets and their followers. The following are a few sample verses that narrate such stories.

*"And We said: "**O Adam**! Dwell you and your wife in the Paradise and eat both of you freely with pleasure and delight of things therein as wherever you will, but come not near this tree or you both will be of the Zalimun (wrong-doers)."*
002:035 *Translated by Khan*

*" (Muslims), say, "We believe in God and what He has revealed to us and to **Abraham, Ishmael, Isaac,** and their descendants, and what was revealed to **Moses, Jesus**, and the Prophets from their Lord. We make no distinction among them and to God we have submitted ourselves."*
002:136 *Translated by Sarwar*

"And We bestowed upon him Ishaque (Isaac) and Ya'qub (Jacob), each of them We guided, and before him, We guided Nuh (Noah), and among his progeny Dawud (David), Sulaiman (Solomon), Ayub (Job), Yusuf (Joseph), Musa (Moses), and Harun (Aaron). Thus do We reward the good-doers."
006:084 *Translated by Khan*

Code of Conduct

The Quran provides a code of Conduct for different aspect of life and society. This encompasses Quranic verses providing instruction on the subjects of privacy, slavery, adoption, social manners, charity, divorce, dowry, extravagance, family, inheritance, mahr, mosque, orphans, waiting period for widows and wills. The named topics are not a complete list of subject areas covered by the Quran but a few examples. A few exemplary Quranic verses focused on by the scholars when giving advice on privacy, Iddah, i.e. prescribed waiting period for widows, and divorcees are given below:

"Believers, your slaves and the immature people must ask your permission three times a day before entering your house: before the morning prayer, at noon time and after the late evening prayer; these are most private times. After your permission has been granted, there is no harm if they come into your presence from time to time. This is how God explains His revelations to you. God is All-knowing and All-wise."

024:058 *Translated by Sarwar*

"O you who believe! when you marry the believing women, then divorce them before you touch them, you have in their case no term which you should reckon; so make some provision for them and send them forth a goodly sending forth."

033:049 *Translated by Shakir*

"Divorce may be retracted twice. The divorced woman shall be allowed to live in the same home amicably, or leave it amicably. It is not lawful for the husband to take back anything

he had given her. However, the couple may fear that they may transgress GOD's law. If there is fear that they may transgress GOD's law, they commit no error if the wife willingly gives back whatever she chooses. These are GOD's laws; do not transgress them. Those who transgress GOD's laws are the unjust."

002:229 *Translated by Rashad*

Islamic Rulings

The Quran includes Quranic verses on Islamic rulings about ablution, abortion, breastfeeding, adultery, chores of business, making contracts, corruption, debt, equality, food, haram (forbidden things), qisas, unmarriageable kin, usury, veiling and the concept of zihar. The following examples relate to Islamic rulings on breastfeeding, the concept of veiling for men and dealing with debt:

"Mothers shall suckle their children for two whole years; (that is) for those who wish to complete the suckling. The duty of feeding and clothing nursing mothers in a seemly manner is upon the father of the child. No-one should be charged beyond his capacity. A mother should not be made to suffer because of her child, nor should he to whom the child is born (be made to suffer) because of his child. And on the (father's) heir is incumbent the like of that (which was incumbent on the father). If they desire to wean the child by mutual consent and (after) consultation, it is no sin for them; and if ye wish to give your children out to nurse, it is no sin for you, provide that ye pay what is due from you in kindness. Observe your duty to Allah, and know that Allah is Seer of what ye do."

002:223 *Translated by Pickthal*

"Tell the believing men to lower their gaze (from looking at forbidden things), and protect their private parts (from illegal sexual acts, etc.). That is purer for them. Verily, Allah is All-Aware of what they do."

024:030 *Translated by Khan*

"O you who believe, when you contract a debt for a fixed time, write it down. And let a scribe write it down between you with fairness; nor should the scribe refuse to write as Allah has taught him, so let him write. And let him who owes the debt dictate, and he should observe his duty to Allah, his Lord, and not diminish anything from it. But if he who owes the debt is unsound in understanding or weak, or (if) he is not able to dictate himself, let his guardian dictate with fairness. And call to witness from among your men two witnesses; but if there are not two men, then one man and two women from among those whom you choose to be witnesses, so that if one of the two errs, the one may remind the other. And the witnesses must not refuse when they are summoned. And be not averse to writing it whether it is small or large along with the time of its falling due. This is more equitable in the sight of Allah and makes testimony surer and the best way to keep away from doubts. But when it is ready merchandise which you give and take among yourselves

> *from hand to hand, there is no blame on you in not writing it down. And have witnesses when you sell one to another. And let no harm be done to the scribe or to the witnesses. And if you do (it), then surely it is a transgression on your part. And keep your duty to Allah. And Allah teaches you. And Allah is Knower of all things."*
>
> *002:282* *Translated by Maulana*

(v) Essential Books for Understanding the Quran

Scholars believe that it is not possible to understand the Quran without background knowledge of Islamic history (Mutahhari 1983). The reason presented is the fact that the Quran was revealed gradually over a long period of time and in fractional divisions of verses. Each verse of the Quran has a historical context attached to it. This context is referred to as the Shan-e-Nuzul of that verse. Knowledge of the Prophet Muhammad's life, acts and sayings is similarly regarded as vital when interpreting the meanings of Quranic verses.

> *"He it is Who sent among the unlettered ones a Messenger (Muhammad SAW) from among themselves, reciting to them His Verses, purifying them (from the filth of disbelief and polytheism), and teaching them the Book (this Qur'an, Islamic laws and Islamic jurisprudence) and Al-Hikmah (As- Sunnah: legal ways, orders, acts of worship, etc. of Prophet Muhammad SAW). And verily, they had been before in mainfest error;"*
> *062:002* *Translated by Khan*

Hadith & Sunnah

Hadith means "saying", and Sunnah refers to the doings of the Prophet Muhammad (PBUH). According to the Qur'an itself, the interpreter of the Qur'an par excellence was the Prophet Muhammad (PBUH) and He was sent as a messenger of Allah. The Qur'an states:

> *"We have revealed to you the Reminder that you may make clear to men what has been revealed to them … "* *016:044*

These terms Hadith and Sunnah are often used interchangeably due to the similarity of the concepts. The Hadith books report the narrations of the companions of the Prophet Muhammad (PBUH) about his sayings and doings. This involves chains of narrators for certain sayings where the narrator refers to the narration given by someone else or comments on the evidence provided by the witness of Sunnah. A collection of six books known as al-Kutub-ul-Sittah is widely accepted as correct knowledge of Hadith and Sunnah. These are referred by most of the sects of Islam. The following list of these books is ordered based on their authenticity as compiled by Hajj Gibril (Schlaefer 2007):

- Sahih Bukhari, collected by Imam Bukhari (d. 256 A.H., 870 C.E.), includes 7275 Hadiths
- Sahih Muslim, collected by Muslim b. al-Hajjaj (d. 261 A.H., 875 C.E.), includes 9200 Hadiths
- Sunan al-Sughra, collected by al-Nasa'i (d. 303 A.H., 915 C.E.)
- Sunan Abu Dawood, collected by Abu Dawood (d. 275 A.H., 888 C.E.)
- Jami al-Tirmidhi, collected by al-Tirmidhi (d. 279 A.H, 892 C.E)
- Sunan ibn Majah, collected by Ibn Majah (d. 273 A.H., 887 C.E.)

Tafsir

Tafsir is the exegesis or interpretation, usually, of the Quran. Mufassiruun (Muslim scholars who have written Tafsir) agree that to produce a legal Tafsir on the Quran it requires mastering skills such as Classical Arabic, Arabic philology, Arabic etymology and Arabic morphology. Besides knowledge and understanding of the linguistic features, dialecticisms of the Quran, Islamic Jurisprudence, Islamic Fiqh, Islamic history and Hadith and Sunnah are also classed as essentials.

There are several collections of Tafsir books compiled by different Mufassiruun. But the commentary Tafsir Ibn-e-Kathir, written by Ismail Ibn-e-Kathir is considered as the most authentic Tafsir of the Quran. He follows a specific methodology in his commentary. Firstly he discusses the Quranic verse under consideration with reference to all the Quranic verses that are *directly* related to that verse. At times this requires cross-references to other verses that do not relate to the actual Quranic verse under the microscope but they are referenced as they might describe some important aspects of the subject under discussion. Secondly, Ismail Ibn-e-Kathir makes references to the Hadith and Sunnah as and where appropriate to support the argument. He includes narrations from collections of authentic and knowledgeable authors, e.g. Ibn-e-Abbass and Ibn-e-Masud.

A recent research work (Muhammad 2013) has made Tafsir Ibn-e-Kathir available for computational research on the Quran. This work has been applied to further research as a source of additional semantic knowledge on the Quran (Jilani 2013).

II – Peculiar Characteristics of the Quranic Text

The language of the Quran exhibits peculiar characteristics. The Quranic text displays a usage of a rich variety of linguistic features. These are uncommon, unusual and significant though not unique to the Quran. Other religious texts present occasional samples of these features but these features are rare in texts collected from sources such as television programs, news wires, conversations, novels and alike. Much knowledge presented in the Quran is encoded via subtle use of words, grammar, allusions, links and cross-references

(Atwell 2010). The following is a brief description of some features relevant to this research project:

(i) Figurative Meanings of Words

The Quran presents the key concepts using *figurative* words instead of words offering the literal meanings. For instance one such key concept discussed repeatedly in the Quran is Jannah or paradise or heaven. According to Islamic eschatology the concept *Jannah* refers to a reward place awaiting the followers of Islam which Muslims will reach after resurrection on Yawmul-Hisab (Day of Judgment) and live forever. The word *Jannah* has been used in its literal meanings "Garden" in the following Quranic verse:

> *"But they turned away; so WE sent against them a devastating flood. And WE gave them, in place of their two excellent gardens, two gardens bearing bitter fruit and containing tamarisk and a few lote-trees."*
> *034:016* *Translated by Sher Ali*

Consider the following Quranic verse. This presents the conceptual meaning of the word *Jannah* referring to "Paradise" not the literal meanings "Garden".

> *"And the Garden is brought near for those who guard against evil -- (it is) not distant."*
> *050:031* *Translated by Maulana*

(ii) Information Distribution in the Quran

Another prominent feature of The Quranic text is that the key concepts or subjects are distributed across multiple verses. It happens that often a concept summarized in one verse is elaborated in another. (Sharaf and Atwell 2012). The Quran indicates this phenomenon in the following Quranic verse:

> *"ALLAH has sent down the Best Discourse - a Book, whose verses are mutually supporting and repeated in diverse forms…"*
> *039:023* *Translated by Sher Ali*

Extending the example the concept of *Jannah* has been discussed across several chapters with different frequency. Qurany (Abbas 2009) keyword search for the conceptual meaning *Heaven* suggest the following statistical facts:

- The concept of *Heaven* has been discussed in 359 verses across 81 chapters in The Quran
- The frequency of reference of the concept Heaven ranges between 1 – 15 references per chapter

The Quran presents information on different aspects of the concept *Heaven* such as the origin or heaven, description of heaven, Allah's authority on heaven and eligibility criteria

for deserving heaven. The following verses exemplify different aspects of *Heaven* as described in the Quran.

> *"Verily your Lord is Allah, who created the heavens and the earth in six days, and is firmly established on the throne (of authority), regulating and governing all things. No intercessor (can plead with Him) except after His leave (hath been obtained). This is Allah your Lord; Him therefore serve ye: will ye not receive admonition?"*
> *010:003* *Translated by Yusuf Ali*

> *"Allah it is Who raised up the heavens without visible supports, then mounted the Throne, and compelled the sun and the moon to be of service, each runneth unto an appointed term; He ordereth the course; He detaileth the revelations, that haply ye may be certain of the meeting with your Lord."*
> *013:002* *Translated by Pickthal*

> *"The gardens which have been promised to the pious have flowing streams, everlasting fruits, and perpetual shade. Such is the blissful end of the pious, but hell fire is the terrible end for the unbelievers."*
>
> *013:035* *Translated by Sarwar*

It is notable that each verse provides information on more than one concept. The Quranic verse 010:003 gives some information on *Heaven* and introduce the concept of *Throne* along with other information. More information is appended to this existing knowledge in later verses 013:002 and 013:035.

(iii) Figurative use of Language

Figurative use of language is a prominent feature of the Quran. A few commonly found figures of speech in the language of the Quran are discussed below:

Metonymy

The Oxford Dictionary (Pearsall and Hanks 1998, Stevenson 2010) for British and World English defines metonymy as "the substitution of the name of an attribute or adjunct for that of the thing meant". For example following Quranic Verse 054:013 describes a ship, ark or boat by providing information about its composition such as "made of planks and nails" instead of naming the boat.

> *And WE carried him upon that which was* ***made of planks and nails****.*
> *054:013* *Translated by Sher Ali*

Metonymy in the Quran is used for a number of purposes. Porting the purpose of a particular metonymy when translating across languages is not always straight forward (Ali, Brakhw et al. 2012). For example the Quranic scholar Yusuf Ali refers to *heaviness or abundance of rain* whereas the same metonymic idea was translated elsewhere using the phrase *loosed heaven,* which fails to convey the true meaning behind the use of the metonymic word.

> *"See they not how many of those before them We did destroy?- generations We had established on the earth, in strength such as We have not given to you - for whom* ***We poured out rain from the skies in abundance,*** *and gave (fertile) streams flowing beneath their (feet): yet for their sins We destroyed them, and raised in their wake fresh generations (to succeed them)."*
> *006:006* *Translated by Yusuf Ali*

Personification

Personification is a linguistic technique where human features are associated with an inanimate object or idea (Cudden 1992), e.g. trees were 'whispering', clouds 'roared'. This technique is usually adopted to facilitate the comprehensibility of a message by using understandable and familiar human characteristics. Personification is used in two ways in the Quranic text. Firstly *Allah* personifies the earth when communicating a message to human beings, e.g. see Quranic verse 084:003. Secondly personification is used to refer to *Allah* Himself e.g. see Quranic verse 003:026.

> *"And when the earth is* ***flattened out****,..."*
> *084:003* *Translated by Yusuf Ali*

> *"Say (O Muhammad SAW): "O Allah!* ***Possessor of the kingdom****, You give the kingdom to whom You will, and You take the kingdom from whom You will, and You endue with honour whom You will, and You humiliate whom You will.* ***In Your Hand*** *is the good. Verily, You are Able to do all things."*
> *003:026* *Translated by Khan*

Hyperbole

Hyperbole is the linguistic device of exaggeration in written or spoken language. Hyperbole when used intends to add effectiveness in communication by use of words that evoke strong feelings towards a certain concept. The point being made is conveyed using hyperbolic words to add intensity to usual plain words. Following three Quranic verses present some examples of how hyperbole has been employed in the Quran.

*"Those who react Our messages and turn away from them haughtily, the doors of heaven will not be opened for them, nor will they enter the Garden **until the camel pass through the eye of the needle**. And thus do We reward the guilty."*
007:040 *Translated by Maulana*

*"When they came upon you from above you and from below you, and when the eyes **grew wild** and the hearts **reached to the throats**, and you were harbouring doubts about Allah."*
033:010 *Translated by Khan*
*"It will be said, "Enter the gates of Hell, wherein you abide **forever**." What a miserable destiny for the arrogant."*
039:072 *Translated by Rashad*

Metaphors and Similes

Metaphor is defined by the Oxford Dictionary (Pearsall and Hanks 1998, Stevenson 2010) of British and World English as "a figure of speech in which a word or phrase is applied to an object or action to which it is not literally applicable". A simile, on the other hand is defined as "a figure of speech involving the comparison of one thing with another thing of a different kind, used to make a description more emphatic or vivid". A simile can be observed when there is an explicit statement declaring two unlike elements to be common, usually done by adding '*as*' or '*like*' to show the relationship, e.g. as brave as a lion. The use of metaphors and similes facilitates the understanding of abstract theories by relating them to more familiar and understandable concepts.

The use of metaphors and similes in various verses of the Quran seems to both indicate and measure the intensity, seriousness and importance of an event or situation. The use of metaphors and similes conjures the images in the human mind, to encompass the full reality and impact of the details of information being conveyed. For example, consider the use of certain words to describe the different stages of the development of a human baby through words which describe its shape like a leech. These words 'leech-like clot' are an example of simile whereas the actual word used in Classical Arabic is metaphorical.

*"O mankind! if ye have a doubt about the Resurrection, (consider) that We created you out of dust, then out of sperm, then out of a **leech-like clot,** then out of a morsel of flesh, partly formed and partly unformed, in order that We may manifest (our power) to you; and We cause whom We will to rest in the wombs for an appointed term, then do We bring you out as babes, then (foster you) that ye may reach your age of full strength; and some of you are called to die, and some are sent back to the feeblest old age, so that they know nothing after having known (much), and (further), thou seest the earth barren and lifeless, but when We pour down rain on it, it is stirred (to life), it swells, and it puts forth every kind of beautiful growth (in pairs)."*

022:005 *Translated by Yusuf Ali*

*"It is He Who has created you from dust then from a sperm-drop, then from a **leech-like clot;** then does he get you out (into the light) as a child: then lets you (grow and) reach your age of full strength; then lets you become old,- though of you there are some who die before;- and lets you reach a Term appointed; in order that ye may learn wisdom."*
040:067 *Translated by Yusuf Ali*

*"Then did he become a **leech-like clot;** then did (Allah) make and fashion (him) in due proportion."*
075:038 *Translated by Yusuf Ali*

The metaphorical use of the word 'leech' not only describes its shape but its characteristics of depending on the mother, sucking nutrients from the blood stream of the mother, etc., just like the parasite leech. The following Quranic verse describes certain events and happenings expected to occur on the *Day of Judgement* to characterise *Doomsday* itself.

"And when the heaven shall be stripped off and taken away from its place; And when Hell-fire shall be kindled to fierce ablaze. And when Paradise shall be brought near,"
081:011 – 13 *Translated by Khan*

(iv) Grammatical Nature of the Quranic Text

The linguistic structure of the Quranic language at times seems to defy the conventional grammatical rules. For example verb-preposition binding as pointed out by Abdul Baquee Muhammad in his PhD thesis (Muhammad 2013), enables Allah to convey multiple meanings through precise and concise choice of words. This also makes the language of the Quran accessible for translation into hundreds of different languages across the world. This significant feature indicates and confirms the universality of the Message of Quran and indicates its importance and significance for the entire human race across cultural, geographical, and linguistic divides.

(v) Mental Schema and the Quran

Every human being is an individual and has a different schema, i.e. set of knowledge and experience, about the world. Similarly every individual has a different way of building their schema. The figurative and linguistic quality of the Quran is such that it uses specific words, phrases and figures of speech as explained in section 2.2. These figurative words in turn trigger different responses in different individuals according to their schemas. Thus, every individual reads or listens to the same words of the one Quran but understands it according to their schema. As different cultures and languages have different ways of effecting,

building and influencing the schema of a person, the Quranic language responds to these different requirements of people across cultural and linguistic boundaries. For example the word *'Jannah'* in the Quran immediately triggers a schema in the mind of an individual of a place with tress, grass, rivers, flowers and wind and shade with benches and even play area for children. Different people will understand *Jannah* according to their own conception of a *garden* they have seen or visited or read about or even how they desire it be. So it can be supposed that different people will enter *Jannah* with different expectation of a *garden* and will find their *Jannah* according to how their schema allows them to imagine it.

(vi) The Quran's Appellation to Human Senses

The human mind is designed to be influenced through referring to and using the human senses. For example, a person is more likely to believe what he can see, hear, feel, touch and taste. A highly significant feature of the Quranic language is that it uses figures of speech to create similar sensations, emotions, and effects on the human mind as a person would feel if they were directly engaged or witnessing the situation (for example – Silk sensation, Heat of fire, Juice of Pomegrates). Furthermore, the Quranic language appeals to the sixth sense of human beings through its poetic and rhythmic construction of language using rhythmic schemes as well phonetics and phonemes which not only make it aesthetically pleasing but the emotional appeal it generates in the listener is often too emotional and difficult to be packed into words.

This clearly indicates that the linguistic structure of the Quran is designed to appeal to the human senses, understanding, mind and imagination. This explains why listening to the recitation of the Quran is often an emotional experiences which has been confessed, experienced and praised by non-Muslims. This again explains the reason why many non-Muslims from the time of the Holy Prophet Muhammad till this very day have accepted Islam by merely listening to its recitation.

(vii) Imagery

A number of research articles present in depth arguments about how and why certain linguistic features have been adopted in the Quran (Rhodora 2012) (Sheikh 2006) (Tzortzis 2008). These linguistic features not only appeal to the human senses but create certain images along with their impact in our minds to fully convey the effect of the message being told.

III – English Translations of the Quran

Muslims hold that the original data format of the Quran was rendered in spoken Classical Arabic (Atwell 2010). Tradition maintains that translating the Quran has been considered

illegitimate at times. Detractors of translating the Quran find support in the second verse of the chapter Al-Yusuf which reads *"We have sent it down as an Arabic Quran".* Baker et al (Baker 2008) emphasise that "if and when used, translation would function merely as a commentary, explaining or paraphrasing the source-text but not replacing it". Hence any attempt at translating the Quran is essentially a form of exegesis, or could be considered as a 'paraphrase' or 'Basic Interpretation'. The true Quran is the one authored in Classical Arabic only (Whissell 2004) and translations are to be used as a study aid to learn and understand the Quran for non-Arabic speakers. Quranic translations thus are not considered *'holy'* in themselves. Hence by implication, no universally recognised legitimate single translated edition of the Quran exists. Every translation has had to qualify for its legitimate value on the basis of the usual measures of accuracy, relevance and stylistic impact of the translation.

It is vital to note that the research behind this thesis uses multiple English translations of the Quran in multiple ways (parallel corpus or in isolation). Therefore the research discussed in this work is one step away from an English translation of the original document, which is in Classical Arabic. The rationale behind using English translation of the Quran as compared to Classical Arabic is to make Quranic knowledge accessible to non-native speakers of Arabic and the majority of non-Arabic Muslim countries, e.g. Pakistan, Malaysia. The following is the list of English translations of the Quran that have been used for this research work:

- Farid-ul-Haq (2002); *Kanz-ul-Eeman.*
- Maulana Muhammad Ali (2002); *The Holy Quran: Arabic Text with English Translation, Commentary and Comprehensive Introduction.*
- Muhammad Muhsin Khan & Muhammad Taqi-ud-Din al-Hilali, (1999); *The Noble Quran.*
- Rashad Khalifa (1985); *Authorized English Translation of the Quran*.
- Mohamedali Habib Shakir's (1982); *The Qur'an.*
- Sheikh Muhammad Sarwar (1981); *The Holy Qur'an: Arabic Text and English Translation.*
- Maulwi Sher-Ali, (1955); *Holy Qur'an: Arabic Text with English Translation.*
- Abdulallah Yusuf Ali, (1934); *The Holy Qur'an Text, Translation and Commentary.*
- Muhammad Marmaduke William Pickthal (1930); *The Meaning of the Glorious Koran.*
- Maulana Muhammad Ali (1917); *The Holy Quran.*

IV – Why Question Answering on the Quran

The Quran is followed by one of the major world religions – Islam. Followers of Islam are known as Muslims. A research manuscript authored at the Department of Computer Science

in Polytechnic University Puerto Rico, USA informs us that the Muslim world population in 2010 was 23.9% of total world population. The report also predicts this population size to increase nearly by another 10% by 2075 (Kettani 2010). An article published in *Time (Jones 2011)* claims that the overall world population will be over 8 billion people by year 2030. Out of this the Muslim world population is expected to be 26.43%. This article also highlights the fact that the Muslim world population was 1.1 billion in 1990 and this is due to double, making 2.2 billion Muslims in the world by 2030. An analysis report published in 2011 by PEW Research on Religion and Public Life indicates that the Muslim world population is expected to increase by 35% in the next 20 years (PEWResearch 2011).

Muslims consider the Quran as the word of God (Allah) and treat it as (the source of divine) law. Susan et al. state that "Acquiring knowledge, ethics, and a moral worldview is a foundation for achieving what the Qur'an requires of every human being — to enjoin what is good, and prevent what is evil" (Jones 2004). Hence Muslims attempt to justify their life in terms of the Quran. The Quran is regarded as universal, comprehensive, and final (Choueiri 2010). Muslims accept the Quran as a fundamental, definitive and authoritative source for all aspects of life. This authority is attested by the Quran itself several times, e.g. verse 002:001, verse 002:176.

The Quran is widely accepted as the most important source of authority which underpins Muslim religious life (Baker 2008). The Quran contains guidance, jurisprudence, morals, and rituals, in addition to some accounts of historical events (Zaidi S. 2012). Apart from other things the Quran presents commandments relating to issues of everyday life such as finance, inheritance, politics, marriage and divorce etc. These commandments and/or their derivations are a major source of Muslim law. The knowledge and wisdom latent in the Quran being fundamental to Muslim religious life, consulting the Quran is necessary.

Muslims across the world regard the Quran as the only authentic and central book of Islam. This is why Muslims refer to the Quran – to learn and gain the knowledge to be a 'follower of Deen-e-Islam'. Muslims study it for multiple reasons. The Quran is studied by its followers in several ways with multiple motivations and intentions. Portions of the Quran are a part of the essential act of Prayer in Islam. Muslims recite a portion of it every day as a source of spiritual joy and strength. Others refer to it intending to find Quranic verses for personal guidance. One of these motivations is to find commandments relating to their day-to-day needs in the Quran.

V - Computing Research using Other Religious Texts

Other religious texts have been used for NLP research. Researchers at the University of Maryland, USA engaged in a project, *The Bible as a Parallel-Corpus* "to acquire and annotate

texts in order to create multilingual corpora for linguistic research, particularly computational linguistics" (Resnik, Olsen et al. 1999). The research was motivated by the fact that "Religious texts are widely available, carefully translated, and appear in a huge variety of languages" (Resnik, Olsen et al. 1999). Although the research mentioned in this paper focused on the Bible, the same is true for the Quran. The Quran has been widely translated into multiple languages and could potentially serve as an open resource for parallel-corpus research.

Just like the Quran, the Bible comprises books/chapters made up of verses. The English in Quran translations, like the English in Bible translations is not like "normal" English text. The language of the Bible is not plain narrative text; rather some verses are poetry and others are 'poetic'. Such language poses specific problems for NLP research as most of NLP research work has been done on newspapers or similar genres. Cynthia Whissell (2004) recommends "one of the most obvious ways to study a text is by examining its vocabulary and, especially, those words that distinguish it from other texts". Her study on the '*emotion and style'* of English used in (a specific) translation of Quran identifies there are 1,143 words that appear at least five times each in Quran translations. Cynthia Whissell (2004) examined results indicated by ʒ test and identified a distinct group of 61 words whose use is significantly higher in Quran translations as compared to everyday English text, i.e. text collected from television programs, news wires, conversations, novels and other like sources.

There has been some on-going (Abdelsalam, Farouk et al. 2013) NLP research work to develop letter-to-sound rules for the scripture of Sikh religion. This work uses the Guru Granth Sahib as a corpus developed using rules adopted from EMILLE corpora. The aim of this work is to finally develop a text-to-speech tool for this religious document. Research has been conducted on multi-word Expression extraction (Kaplan 2001) on the religious book of Hinduism. This work investigates the role of Multi-word expressions for automated text processing. Daniel Stein evaluated his research using a Spanish version of the Bhagavad Gita.

Section Summary

The above discussion argues that the peculiar characteristics of the Quranic language enable the implicit message to transcend the barriers and boundaries of time and space across cultures and languages. It may also be safe to assert that the language of the Quran is designed as a code that people decipher according to their own abilities and requirements. These features make the language of the Quran universal and able to survive shifting and dynamically changing global circumstances.

Thus the Quranic language not only serves the purpose of education, moral training, imparting information, historical referencing and guidance but also serves as a means of emotional catharsis, spiritual healing and aesthetic pleasure which is pure and devoid of lewdness. It may not be wrong to claim that, therefore, the language of the Quran is designed to last forever till the very last day of the human race. This can and does explain the challenge that Allah has given mankind to produce a similar discourse as soothing, pleasing and complete as the Quranic language.

The above discussion also explains that the preservation of the Quranic message lies in the structure of its language, the strength and rhythmic construction of not only the words but sentences and chapters. Thus, the cohesion and coherence found within the words, sentences and the chapters of the Quran binds and holds together the implicit and explicit message for humanity. This is the reason why Allah asks mankind to 'read' or 'to search' or to 'explore' not only the Quran but knowledge in the universe (Ali, Brakhw et al. 2012). Susan et al indicate that "It is evident that both the obligation to be educated, and the moral, intellectual and cultural concepts of an education in the Muslim tradition are not far removed from similar goals and concepts associated with Western traditions and aspects of education" (Jones 2004). The Quran, being the core of Islamic knowledge, governs the day-to-day life of a Muslims. Hence they refer to it often for education, guidance and learn latent Quranic knowledge.

Lastly, it is important to note that the complex features of the Quran discussed above are also reflected in the translations of the Quran. Hence research on the corpus of English translations of the Quran cannot be expected to produce the same quality results when tested with NLP tools and techniques produced by prior research with English. The need for information from this corpus is very different from the need for information expected from other corpora. Questions asked by users are mostly complex questions which cannot be dealt with as simple factoid questions. This suggests a need to adapt the existing research tools and techniques to this corpus. The next sections discuss the computing research being conducted to achieve this purpose.

2.3 – Computational Research on the Quran

This section presents a review of the computational research conducted on the Quran. Another objective of this chapter is to present a sample of recent and related research work that has informed the research behind this thesis.

I – Cluster Analysis for Automated Classification of Quranic Chapters

As discussed earlier in section 2.2.2, the information and knowledge presented in the Quran is distributed across different chapters and verses. There has been some work on understanding this thematic structure of the Quran in order to find the relationship between different chapters in the Quran (Bewley 1999). This research was aimed at finding an answer to the assertion of whether or not it is possible to classify the chapters of the Quran computationally, based on the semantic content presentation across chapters. It was conducted using transliteration of the Quran and not the Quran in its full Arabic form. This work applied cluster analysis to a large matrix that contained the chapters of the Quran on one axis and a shortened list of unique content words on the other. The shortened list was obtained after eliminating function words and less informative words from the full list of corpus words.

Naglaa Thabet's work (2004 – 2006) also produced a purpose-built stemmer(Sultan, Azman et al. 2011) based on the light-stemming technique. This stemmer aimed to group morphological variants of each word in transliterated form. This work faced issues around sparseness of data in the matrix due to extreme variation in chapter lengths. Thabet attempted to tackle this problem by truncating the analysis initially to the 24 longest chapters of the Quran chosen at random. These had more than 1,000 content words after stemming and elimination of function words. The following table, 2.1, lists the chapters which were selected for the initial experiments. This was further reduced to 500 words in later experiments. Thabet conducted further experiments on this dataset including hierarchical clustering algorithms for lexical frequency data in order to achieve classification of the Quranic chapters, multidimensional scaling, and nonlinear methods to deal with the nonlinearities in her dataset (Rahim 2007). The results of these experiments were informative but compromised due to variation in chapter lengths. This problem reflected the complex nature of the Quran in terms of its content organisation.

Hermann Moisl (Sultan 2011) analysed Naglaa Thabet's work and proposed to introduce a minimum threshold limit and selection of lexical variables connected to the set threshold as a solution to the problem degrading Thabet's experimental results. His proposed solution was based on techniques borrowed from sophisticated statistical sampling distribution. Moisl reduced the sample size for his experiments to only those Quranic chapters that have around 300 content words. Moreover he used only 9 words as analysis parameters which is very minimal in relation to the overall word length of the Quran.

Chapter #	Chapter Name	Content Words	Chapter #	Chapter Name	Content Words
02	Al-Baqarah	5739	24	Al-Nur	1236
03	Al-Imran	3316	26	Al-Shu'raa	1208
04	Al-Nisa	3543	09	Al-Tawba	2345
05	Al-Maidah	2681	10	Yunus	1732
06	Al-Ann'am	2895	11	Hud	1809
07	Al-A'raf	3127	12	Yusuf	1665
08	Al-Anfal	1156	16	Al-Nahl	1729
17	Al-Israa	1464	27	Al-Naml	1069
18	Al-Kahf	1489	28	Al-Qasas	1332
20	Ta-Ha	1265	33	Al-Ahzab	1239
21	Al-Anbiyaa	1077	39	Al-Zumr	1107
22	Al-Hajj	1195	40	Ghafir	1156

Table 2.1: The Quranic Chapters with more than 1,000 words used for Cluster Analysis (Mustapha 2009)

Abdul Baquee Sharaf (Muhammad 2013) argues that Thabet and Moisl's work demonstrated a useful application of clustering algorithms. However, as the work involved only limited portion of the Quran, the results might not be a true projection of other chapters of the Quran. He also believes that the statistical methods used in these researches could have yielded enhanced results if linguistic and domain specific knowledge had been embedded within the data used for this work.

II – Visualisation System for the Quran

There has been recent research in connection to the peculiar features of the Quranic text discussed in the previous chapter. One of these research works, conducted by Aida Mustapha (Mustapha 2009) presented a prototype visualisation system AQILAH for the Quran. The underlying idea for the Quranic Visualisation System is that due to the nature of the Quranic verses, it is impossible to comprehend the meanings of a verse in isolation. It is essential to interpret it within the context of the chapter it belongs to in order to maximise understanding.

In her paper "*Dialogue-Based Visualisation for Quranic Text*", Aida Mustapha states that in order to enable a visualisation system to convey the Quranic text and content with clarity, precision and efficiency such systems must be able to capture co-occurrences of concepts

across chapters in the original textual form and not in graphical or diagrammatic form. The architecture of AQILAH is presented in the following block diagram.

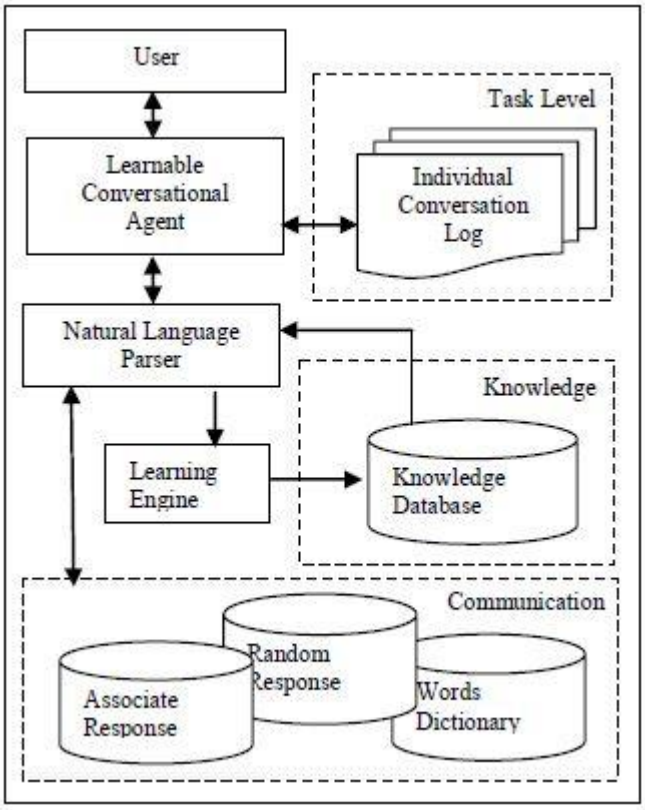


Figure 2.14: Block Diagram showing Architecture of AQILAH Visualisation System (Mustapha 2009)

"Visualization systems supporting the reading of Arabic document for non-Arabic speakers" (Yusof, Zainuddin et al. 2009), presents visualisation systems as a solution to improve reading comprehension of the Quran. It is reasonable to expect to understand a document while one is reading it. But the authors argue that the Quran is a unique document, which unlike usual documents, is often not understood when it is being read by non-Arabic speakers. Thus, this research work discusses the issues while reading this special Arabic document, the Quran. This discussion takes into account visualisation theories and techniques around reading issues. This work also provides a comparative analysis of existing prototype systems offering a solution to these issues.

The work of Yusof et al (2009) also presents an ontology that focuses on the structure of the content of the Quran. Incorporating this ontology enables improved mapping of the knowledge base. Finer details of this work are not available as it has restrictions on access.

III – Quranic Ontologies

The structure of the Quranic corpus is suitable for ontological forms of knowledge representation. Definitions of ontology insist that an ontology specifically represents common, shared conceptual structures (Brewster and O'Hara 2004). Hence an ontology can be seen as a network of concepts and relationships between these concepts. This is valid for any corpus regardless of language. Recent computational research on the Quran has produced an ontology of Quranic concepts (Dukes 2011) representing key topics in the Quran and their relationships (Dukes and Atwell 2011), QurSim: an ontology presenting semantic relationships amongst Quranic verses Sharaf and Atwell (2012) Abdul-Baquee (2013) and QurAna: a corpus of the Quran with pronoun references Sharaf and Atwell (2012). Other ontologies include Mustapha (2009), Yusof, Zainuddin et al. (2009) whereas online Islamic websites provide verse classification based on topics that can be realized as an ontology Salhey (2013) Al-Munajjid (2013) Ali (2013) as each of these is a network or concepts and relations. Most of these resources cover the Quran both in Arabic and English.

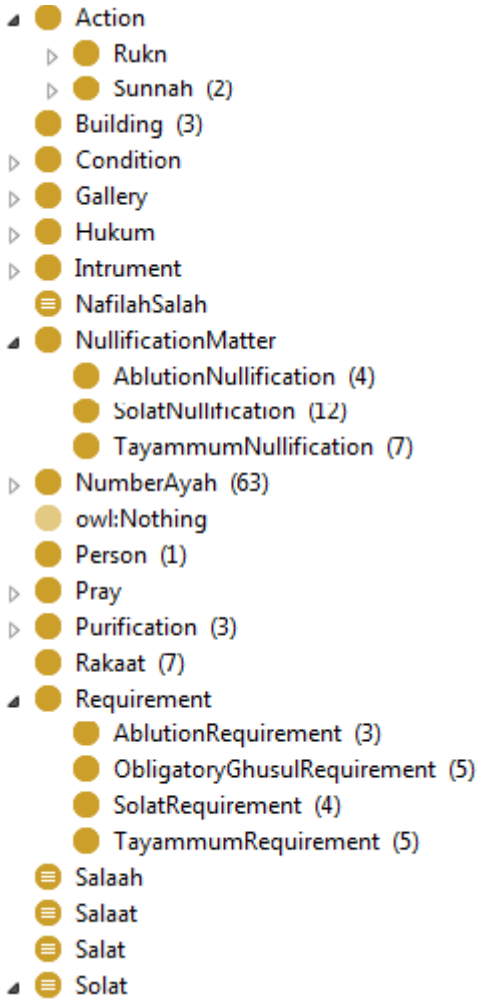


Figure 2.15: Different Levels of Solat Taxonomy (Saad, Salim et al. 2011)

Saidah Saad et al present a discussion on their experience in developing a domain-specific ontology (Saad, Salim et al. 2011). The authors chose Islamic Knowledge as a domain of interest and set the concept of 'Solat' as the focus of this research. As a prototype the knowledge available in the Quran alone has been considered. This is because the Quran is the core source of Islamic knowledge. All other sources derive from the Quran. The authors used Quranic indexes to identify concepts of Solat. Saad et al indicate different levels of the Solat Ontology as shown in the above Figure (2.15).

Online Islamic websites provide verse classification based on topics that can be realized as an ontology (Salhey 2013) (Al-Munajjid 2013) (Ali 2013) as each of these is a network or concepts and relations.

IV – Islamic Question-Answering Systems

There are many online websites that disseminate Islamic information and provide answers to questions. It is important to note that these are not automated systems but involve a layer of human interaction. These systems are a part of the online Islamic community Services. The Q&A system is embedded into other facilities made available to the virtual community. Registered members and guest users are allowed to ask questions using basic interactive web interfaces and Islamic scholars answer these questions offline. These scholarly responses are once again made available via web interfaces as a reply to the inquirer as well as information Q&A pair for new consultations.

There are several community websites available online but one needs to take precautions in selection amongst these websites. There are reference websites that enlist criteria to evaluate Islamic websites as well as discussion about how to verify the information these provide, e.g. Anon (2002). As the focus of this thesis is on Natural Language Processing – Question Answering hence only those websites were researched in detail that disseminated Islamic information in some form of Q&A service using English as the medium of communication. There are several Question Answering repositories available online where scholars (Imams, Muftis) and other domain experts provide answers to user questions asked by real world users.

The following are examples of a few such websites; Islam Question Answering (Al-Munajjid 2013), Study the Quran (Ali 2013), Islam City (Imam 2013), Quran Al-Islam (Salhey 2013), Islam Web (Mufti 2013), Qibla Answers (Keller 2013), On Islam (Scholars 2013), Understanding Islam (Amjad 2001), Alim (Shah 2010), Studying Islam (Rahim 2007), and Renaissance Journal (Sheikh 2006). A brief introduction to these websites has been provided for reference.

2.3.1 Islam Q&A Website

Islam QA (Al-Munajjid 2013) is an authentic Islamic website launched in 1997, maintained and supervised by the scholar Shaykh Muhammad Saalih al-Munajjid. The website provides authoritative answers to questions and counselling advice on personal or social questions asked by anyone – Muslim or Non-Muslim. All materials are approved through formal procedures before publication.

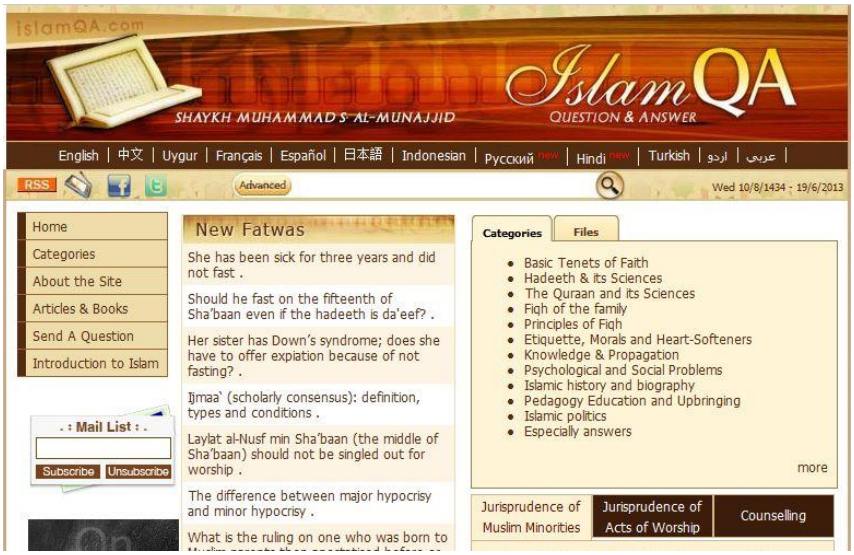


Figure 2.16: Islam Question Answering (1997)

Mostly questions asked by website users outline details of personal situations and seek advice in that context. The scholar Shaykh Muhammad Saalih al-Munajjid and other scholars answer these questions in the form of free conversation and commentary. The PARMS Model Question Set (further details in section 4.1) includes 33 questions from this resource.

2.3.2 Study the Quran Website

Study the Quran (Ali 2013) is an initiative of the Institute of Studies of the Quran established in 2003 with the objective of motivating people to understand the Quran and to be able to live life in accordance to the true spirit of Islam. The organisation aims to provide a multi-lingual platform where people can read and understand the Quran in their own language. The brains behind the project believe that this is essential for people to be able to get to the in-depth meaning of the Quran. Alongside other services and facilities, the website also provides a series of question/answer pairs where answers are provided directly using the Quranic verses without any addition of scholarly comments or references to other Islamic resources. The PARMS Model Question Set (further details in section 4.1) includes 42 questions from this resource.

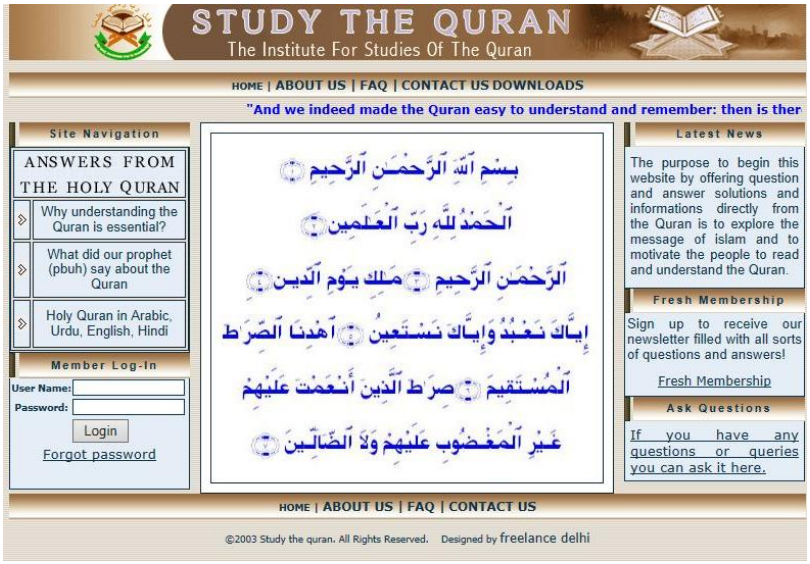


Figure 2.17: Study the Quran Website (2003)

Verses are presented in the translation of the Quranic verse done by the scholar Yusuf Ali (see section 2.2). The website uses the English translation The Noble Quran published in Norwich, UK. This research work is available online (Christmann 2002) and also in hard book form (Bewley 1999).

2.3.3 IslamiCity

IslamiCity (Imam 2013) website launched in 1995 is operated by Human Assistance & Development International (HADI). HADI was incorporated in the State of California in 1991 as a public benefit, charitable, educational and scientific organisation.

Figure 2.18: IslamiCity Website (1995)

The IslamiCity website provides services delivered in a virtual metaphor around community building. IslamiCity has become a popular Islamic information dissemination source and one of the largest e-communities serving a huge number of requests each year.

The IslamiCity website reports over 1,209,471,094 requests serviced since January 2001. This website attracts millions of guests and wide media coverage each year. The PARMS Model Question Set includes 12 questions from this resource. Alongside other services, the IslamiCity website also provides a facility to ask questions of scholars. Question Answer pairs asked by website users are made available for general public reference. But these are processed to conform to confidentiality measures. These QA pairs are classified according to topics and grouped in alphabetical order for quick retrieval.

2.3.4 Islam Web

Islam Web (Mufti 2013) launched in 1998 was designed to enrich knowledge and appreciation of Islam. The mission statement of the website is to increase general public awareness of varied concepts of Islam. The website audience are Muslims and Non-Muslims alike, formal and casual visitors, knowledgeable users and new converts.

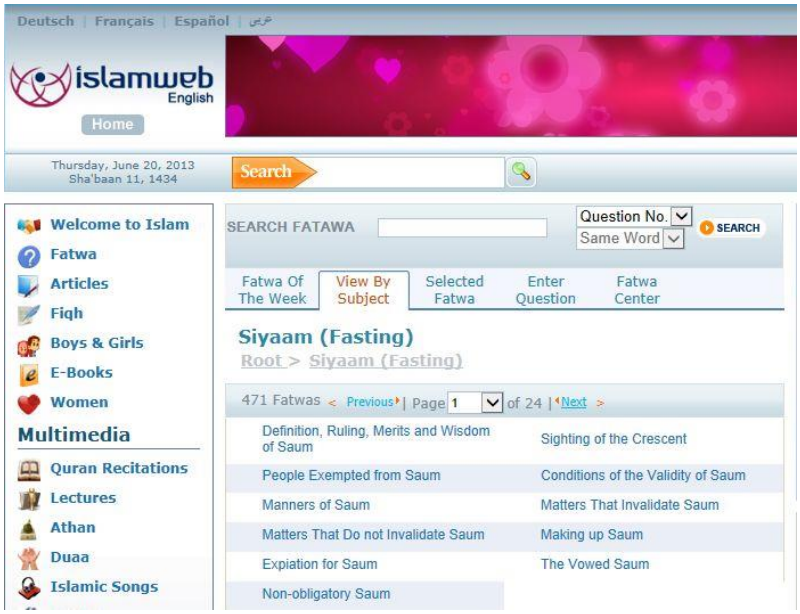


Figure 2.19: Islam Web Website (1998)

The information disseminated by Islam Web encompasses basic tenets of Islam, Islamic law, traditions and the practices of the Prophet Muhammad. Islam Web provides scholarly answers (*Fatwas)* on questions asked by website users, besides other services such as Islamic multimedia, Islamic calendar, and prayer and fasting timetables. Fatwas issued by the Mufti (scholar) include references from the Quran and other key Islamic books of Hadith

and Tafsir. PARMS Model Question Set (further details in section 4.1) includes 2 questions from this resource.

2.3.5 On Islam

The OnIslam (Scholars 2013) website emerged as an editorially independent project launched in 2010 after success of the IslamOnline.net project initiated in 2000. This website is an umbrella effort of Mada Media Development Association (MMDA), a non-profit Egyptian organisation with a global outlook.

Figure 2.20: OnIslam Website (2010)

The main website components are news, Islamic information literature, human development events, and counselling and scholarly advice. Generally it provides information on different aspects of Islam and issues around practicing the Muslim life. The aim of this website is to provide theory and practical support by counselling through answering the real questions of people in need. The website claims to provide credible answers to questions and counselling advice by ensuring professional conduct, scientific accuracy and accurate reporting of Islamic sources of information. The PARMS Model Question Set (further details in section 4.1) includes only 1 question from this resource.

2.3.6 Qibla Answers

The Qibla Answers (Keller 2013) website is managed by a team based in the USA and Australia. The aim of the Qibla Answers scholarly team is to provide authoritative guidance on practical aspects of Islamic knowledge. The website service follows the beliefs of the Sunni sect of Islam. The questions answered by the website are classified in multiple categories depending on the concepts under consideration. Each category further presents

answers groups in descending order of question date. The website also records the number of guest visits for each question. The PARMS Model Question Set (further details in section 4.1) includes only 1 question from this resource.

18143	Please tell me what to do...I can't stop my sins..	39321	12/29/2007
18088	Hatred towards father	9216	12/24/2007
18000	Difference in way of life in our marriage	31946	12/16/2007
17409	Can I tell my five year old daughter that Allah is her best friend?	15189	10/29/2007
17400	Husband refuses to pray	9181	10/29/2007
17239	Women, Work, and Circumstances at Work	16197	10/16/2007

Figure 2.21: Qibla Answer Website group of questions (2010)

2.3.7 Understanding Islam

The Understanding Islam website was an initiative that started as a free server-based webpage back in 1997 and later migrated to a registered domain in 1999. This was later improved to its current version which has been available since May 2001.

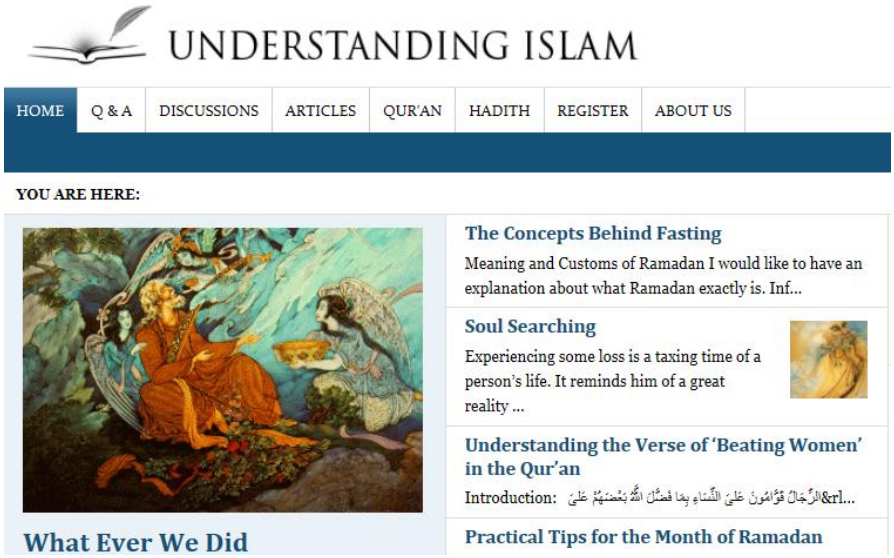


Figure 2.22: Understanding Islam Website (2001)

The motivation behind this website was to firstly to contribute to the available resources on Islam and also provide counterpart information in reply to criticisms made of the Quran and Islam by other religious websites. The service objectives are to present unbiased Islamic knowledge, answer user questions about Islam and disseminate Islamic information in order to reply to criticisms of Islam as well as contribute to the cause of Da'wah (propagation of

Islamic thought). The website also hosts a sister website: Islamic Students – Research Association (ISRA) since 2011.

The website links several external technical resources that make forums available for Question Answering and discussion. These also make it easier for people to take scholarly advice on personal problems but there is not much evidence found about the scholarly levels. So PARMS Model QA set [QID#26 - Appendix 4] adopts only one user question from this website.

2.3.8 Alim

Alim.org is an initiative of Alim Foundation, a non-profit organisation. The origin of this website is connected to Alim CD-ROM, a software development project aimed at making Islamic knowledge available and accessible in a new medium for research and study. The organisation applies technology for the betterment of Islamic information dissemination services. The purpose of this service as indicated is to provide an online educational tool, virtual social networking web-space and a general-purpose web-portal.

Figure 2.23: Understanding Islam Website (2010)

This website has been rigorously used in the process of designing the PARMS QA Set (further details in section 4.1) but it does not have any questions included from this website. This website provides an Index-based search facility for the Quran. This feature has proved very useful in the design process along with the Concept index of the Quran. The details of the design process have been discussed in Chapter 4.

2.3.9 Quran Al-Islam

Quran Al-Islam.com is an Islamic website providing services in the area of the Quran and Hadith. The website provides online software tools such as Inheritance calculator to apply Islamic rulings to practice.

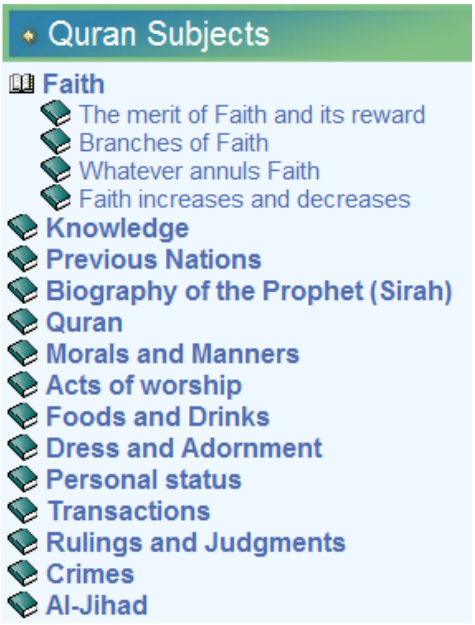


Figure 2.24: QA Categories by Quran Al-Islam.com Website (Salhey 2013)

The Quran Subject feature on Quran Al-Islam.com has been used to assist the design process of the PARMS Model QA set (further details in section 4.1).

2.3.10 Studying Islam

Studying Islam is an affiliate of Al-Mawrid, an online Islamic teaching resource and service.

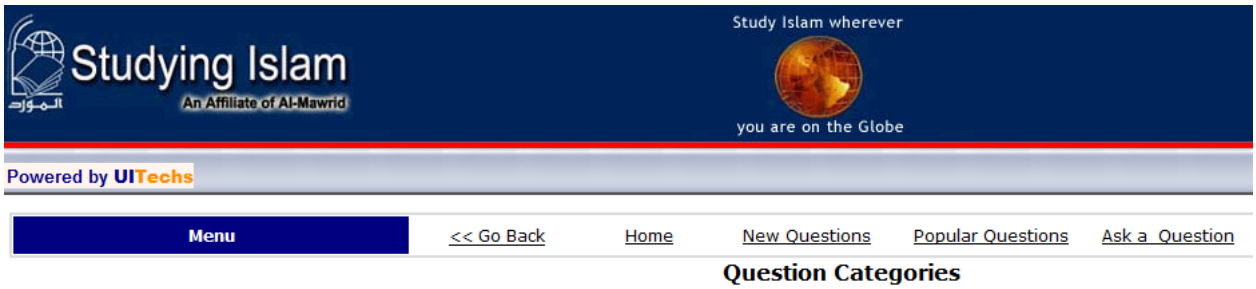


Figure 2.25: Studying Islam (2007) Website

PARMS uses a Question Categories feature in connection with above listed websites to identify related question and verses for popular Islamic topics.

V – Quranic Verse Retrieval Problem

Information Retrieval from the Quran has been a popular research topic for the past few decades. The problem of retrieving information in the form of relevant Quranic verses for a particular topic is also known as the *Quranic Verse Retrieval Problem*. There have been several attempts to address the core issues around the problem area.

Muhammad Fauzan et al (Noordin and Othman 2006) conducted a survey on the structure and linkages of 125 websites that offer access to Quranic Texts. The research findings of their work *An Information System for Quranic Texts: A Proposed System Design* reveals that these websites lack a standard structure. They found that websites usually disseminate Quranic information in the form of Quranic texts, translations, recitations and relevant excerpts from Tafsir. They also provide links to other related websites. Fauzan et al believe the following diagram depicts the best hierarchical structure of Islamic sources as proposed by previous research. They believe this structure should be followed by these websites:

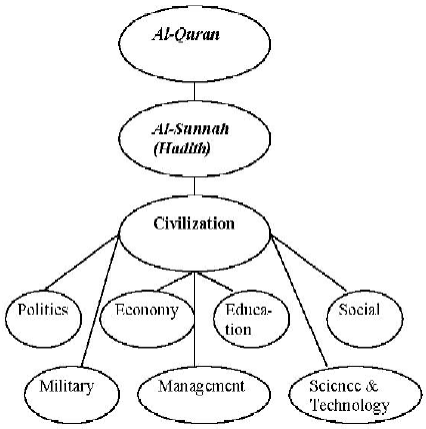


Figure 2.26: Hierarchical Structure of Islamic sources

There has also been research in cross-language Information retrieval using the Quran as a case study. A recent research (Mohd Yunus 2008) investigates a dictionary-based approach to Cross Language Information Retrieval (CLIR) using the Quran. This study involves Malay-English and English-Malay translations using dictionaries. This work uses stemmers for both the English and Malay languages. This work suggests that "correct identification and translation of Multi-word terminology is the single most important process in CLIR".

Another work *Search Engine System for the Holy Quran* (Ismail 2009) involves research on a keyword based search engine especially for the Quran. It focuses on the correction of spelling mistakes caused usually by using vernacular languages or Arabic dialect words. This research work proposes a query correction mechanism that converts such mistaken words

in Al Rasm Al Othmani which is the Arabic text in the Quran. Besides, this work offers a prototype implementation which uses an enhanced version of the existing ISRI stemming Algorithm in order to search Al Rasm Al Othmani Arabic text from the Quran. Details of the proposed query correction method and the prototype system have not been made available for public use as the thesis is restricted to internal University use only.

Recent evaluations of the *Verse retrieval problem* have been made (Sultan 2011; Sultan, Azman et al. 2011). Sultan et al conducted experiments to evaluate the performance of Terrier (Plachouras, He et al. 2004), which is an existing state of the art IR system developed by Glasgow University. This work involved the creation of a query set in Malay and English for a pre-defined compilation of indexed topics in the Quran. It relies on Google Translate for translation between Malay and English query terms but this compromised the quality of results. The authors propose to use a bi-lingual dictionary in order to achieve better cross-language IR results. The research also shows that the IR system investigated performs poorly when applied to the Quranic Verse Retrieval Problem. This perhaps is due to the complex nature of the Quran itself as discussed in the previous chapter. Further works on Quranic Information Retrieval include Qurany Tool (Abbas 2009) ,which is an online keyword-based Quranic search facility. IR research using specialised stemmers for the Quran includes *Semantic Query with Stemmer for Quran Document Results*, a work by Yunus, Zainuddin et al. (2010).

A more recent research work conducted at Alexandria University (Abdelsalam, Farouk et al. 2013) attempts to build a comprehensive system for Arabic Question Answering for the Quranic text which uses semantic methods to develop understanding of the natural language question. This work aims at developing an IR module capable of dealing with the verse retrieval problem semantically, a question classifier using the SVM classification technique, and finally use a ranking method for identifying the best answer to a given question. It also claims to incorporate Tafsir for the presented answer. But the description of the system architecture is unclear and no methodical approach has been adopted for evaluation of the results. There is only a mention of using 54 questions for evaluation purposes but there are no details on the evaluation system/framework used for this work.

The KISS project at Sheffield University (Sanderson and Alberair 2001), AIR project at Syracuse University and QARAB (Hammo, Abu-Salem et al. 2002), a question answering system to support the Arabic language are a few works that do not involve the Quran but have contributed knowledge to IR works involving the Quran.

VI – Quranic Corpus

There also has been work on making the Quran available in the form of a corpus for computational research. The Quranic Corpus works studied in connection to PARMS include the Eight-Parallel Quran by Clay Smith (Smith 2009), Semantic Quran (Sherif and Ngomo 2009) and Quran Corpus for Juz Amma (Aisa Mustapha 2012).

Eight-Parallel Quran contains eight English translations of the Quran presented by different authors compiled in parallel. Eight-Parallel Quran is compiled and copyrighted by Clay Smith for academic and learning purposes. It is available free but must be used in its 'actual' compiled form. Explicit author permissions are required otherwise. The "Parallel Quran" translations Smith C. (2009) present multiple translations for each verse of the Quran in parallel, i.e. all translations of verse one are presented and then all translations of verse two are presented and so on. The latest version of the corpus includes eight translations where each translation belongs to a unique author. The corpus includes translations made between 1917 and 1977 by reliable translators. Some translators use 'Biblical' English (Pickthal, e.g. 'shall ye wisheth') while others use modern English.

107:002 Khan	*That is he who repulses the orphan (harshly),*
107:002 Maulana	*That is the one who is rough to the orphan,*
107:002 Pickthal	*That is he who repelleth the orphan,*
107:002 Rashad	*That is the one who mistreats the orphans.*
107:002 Sarwar	*It is he who turns down the orphans*
107:002 Shakir	*That is the one who treats the orphan with harshness,*
107:002 Sherali	*That is he who drives away the orphan,*
107:002 Yusufali	*Then such is the (man) who repulses the orphan (with harshness),*

Figure 2.27 Sample from Eight-Parallel Quran

The above figure presents a sample verse (107:002) from the Parallel Quran. The first 3 digits (before the colon sign) assigned to each translation present the number of the chapter and the next 3 present the number of verse in that chapter. The above is the second verse of Chapter 107 of the Quran. The name of the author is appended at the end of each number sequence. Lastly the translation of the verse is presented.

Semantic Quran provides a translation of the Quran in forty-three different languages. Lastly, Juz Amma is a bilingual knowledge rich Quranic corpus. It uses an English – Malay language pair. The work includes the Quranic verses as well as offering additional knowledge from Tafsir and Hadith for every verse. The scope of this work is initially limited to only the last Juz' out of thirty Juz' of the Quran, but it could prove a useful corpus for

limited experiments. This work is copyrighted and has not been made available for public research.

VII – Linguistic Work on the Quran

Some prominent works in this aspect of computing research on the Quran include the listing of key concepts described in the Quran presented at the Text Mining the Quran website (Abdul-Baquee 2013) and research thesis (Muhammad 2013), QurAna: Pronoun Resolution work for the Quran (Sharaf and Atwell 2012), QurSim: Verse similarity work adapted from Ibn-Kathir (Sharaf and Atwell 2012) and linguistic annotation work on classical Quran corpus (Dukes 2011). QurAna and QurSim are discussed further in this chapter below.

The research work by Abobakar Ali et al reports on the linguistic difficulties faced by authors while translating the Quran (Ali, Brakhw et al. 2012). Haja's work deals with the current discrepancies between ETL dictionaries in capturing Islamic terms (Mohideen 2007). Haja classifies the distortions of Islamic terms in three areas, namely misspellings, mispronunciations and mistranslations when Arabic words are translated into English. These difficulties and limitations are imported into all translations of the Quran. Hence the same linguistic characteristics contributed to the complexity of the underlying corpus used by the PARMS QA System. A research article *Assessing English Translations of the Qur'an* (Mohammed 2005) provides some useful information in connection to identifying the best available translations of the Quran.

VIII – QurAna

The Quran is characterized by very frequent use of anaphors (Sharaf and Atwell 2012) and the majority of anaphoric devices in the Quran involve pronominal anaphora. There is a need for a language resource which incorporates the ability to resolve pronoun antecedents. QurAna is a large corpus created using the Arabic Quranic text where personal pronouns used in each Quranic verse have been replaced by their personal pronoun antecedents. Abdul-Baquee Muhammad Sharaf explains that this characteristic is partly due to the morphology of the Arabic language itself. He also explains that QurAna as a language resource has wider applications in NLP such as machine learning, information extraction and question answering. As a sample some visualizations of machine learning applications using QurAna have been presented on the author's wiki page (Abdul-Baquee 2013). The author explains the peculiar characteristics of the Quran by presenting Quranic verse 024:031 as an example:

> *And tell the believing women to lower their gaze and be modest, and to display of their adornment only that which is apparent, and to draw their veils over their bosoms, and not to reveal their adornment save to their own husbands or fathers or husbands' fathers, or their sons or their husbands' sons, or their brothers or their brothers' sons or sisters' sons, or their women, or their slaves, or male attendants who lack vigour, or children who know naught of women's nakedness. And let them not stamp their feet so as to reveal what they hide of their adornment ..*

As explained by the author, the Arabic version of this Quranic verse contains 27 pronouns but these have been reduced to 16 when translated into English by Pickthall. There is a need to resolve these pronouns to set the context of the referent for each pronoun. Besides this helps deal with ambiguity also. Unless resolved it is not possible to identify which entity a particular pronouns relates to. This can be observed from following two Quranic verses:

> *And when We made a covenant with you and caused the mount to tower above you [2:63]*

> *And when We shook the Mount above them [7:171]*

The author considers the above examples where a particular concept 'Children of Israel' has been referred to within the Quranic text at a particular point and the pronouns in later Quranic verses make reference to the concept. This pattern is very frequently observed in the Quranic text where the reader is required to apply personal understanding and intuition as there is often no explicit mention of an antecedent reference. It is vital to disambiguate the pronoun references in order to understand the true meaning of the Quran.

QurAna corpus is a novel attempt to address the above stated issues for the Quranic text. It presents the antecedents of the Quranic pronouns in the form of an ontological list of concepts. The annotation scheme and annotation process have been described in the research paper *QurAna: Corpus of the Quran annotated with Pronominal Anaphors* (Sharaf and Atwell 2012) and further details on this work are available in the research thesis report (Muhammad 2013). Abdul-Baquee Muhammad Sharaf claims that QurAna has interesting computational linguistic applications and will eventually motivate and attract more research in a new direction involving the Quran. As a consequence, the QurAna ontology has been used in further research work involving the Quran as a case study (Jilani 2013). Details of how QurAna has been benefited from as a source of additional semantic knowledge for the Quranic text are described in chapter 5, section 5.3.

IX – QurSim

The Quran is a complex knowledge source as it has much knowledge encoded via subtle use of words, grammar, allusions, links and cross-references (Atwell 2010). Key concepts or subjects in the Quran are distributed across multiple verses in the Quranic text. Each instance of the concept adds more information and meaning to the prevailing context. Hence it is vital to take all instances of the concept into consideration for understanding the overall meaning. This characteristic of the Quran has been discussed in detail in the section 2.2. The information distribution property of the Quran makes it attractive for computational research on semantic relatedness in natural language in short texts. Abdul-Baquee Muhammad Sharaf has created such a dataset using the original Quranic text called QurSim (Sharaf and Atwell 2012).

QurSim is a large corpus that enlists semantically related verses as ontological links. The relatedness and similarity measures have been derived from reliable sources of domain knowledge. It uses the scholarly work, Tafsir-Ibn-Kathir, originally compiled by well-known Muslim Scholar Ismail Ibn Kathir who died in 1373 CE. His work is sometimes considered the gold standard commentary of the Quran. Besides literary commentary on the Quran, Tafsir-Ibn-Kathir also provides information on similar verses in the Quran. More information on this Tafsir has been given in earlier section 2.2.

Primary information on objectives, compilation process, design methodology, applications and challenges met in putting this corpus together have been published in *QurSim: A corpus for evaluation of relatedness in short texts* (Sharaf and Atwell 2012) and further details have been reported in the thesis (Muhammad 2013). Online visualizations of experimental results are provided on the author's wiki page (Abdul-Baquee 2013). This includes verse relatedness tables, a concept cloud, Quranic verse networks and chapter relatives for all Quranic Chapters.

QurSim can perhaps be seen as a gold-standard language resource for computational research involving the Quranic Text. Abdul-Baquee proposes its application in areas such as word sense disambiguation, information extraction and retrieval, automatic indexing, lexical selection, text summarization, automatic correction of errors, and word and text clustering. QurSim is a novel application in the context of Question Answering (Jilani 2013) using the Quran as a case study. This work utilises this resource for Answer Re-ranking. The assumption behind this application is that in a list of candidate answers for a given question, semantic similarity between the Quranic verses should be considered to improve the rank of a particular answer. Details of the re-ranking algorithm are presented in this thesis (section 3.2.4).

X – QurTerms

The original text of the Quran uses Classical Arabic but it has been translated into several other languages including English. English is considerably regarded as the most important language in science (Ammon 2001, Kaplan 2001, Caplan 2013). This has motivated computational research using English translations of the Quran (Mohd Yunus 2008, Yunus, Zainuddin et al. 2010, Yunus, Zainuddin et al. 2010, Sultan, Azman et al. 2011, Jilani 2013). Quranic linguistic resources are available in English e.g. Concept ontology (Abbas 2009), word-by-word translation of the Quran (Dukes 2011), QurAna (Sharaf and Atwell 2012), QurSim (Sharaf and Atwell 2012) and corpus of English translations of the Quran (Sherif and Ngomo 2009, Smith 2009) but there is a greater need for more dedicated NLP resources for English translations of the Quran. Besides there is a need to access Quranic knowledge in English due to the huge number of non-Arabic speaking Muslims. This has been discussed earlier in section 2.2.

One approach to satisfy this need could be to use existing English NLP tools and techniques, e.g. PoS taggers, Stemmers, WordNet (Miller, Beckwith et al. 1990, Miller 1995, Fellbaum 1998, Pedersen, Patwardhan et al. 2004). There has been some mention of adopting existing Stemmers to suit the Quranic text in cross language works done for the English-Malay pair with application to the Quran (Mohd Yunus 2008, Yunus, Zainuddin et al. 2010, Yunus, Zainuddin et al. 2011). WordNet (Pedersen, Patwardhan et al. 2004) has proven value in NLP research (Morato, Marzal et al. 2004) but it is not suitable for the Islamic terms that have been imported into English language.

The Quran, being the core of Islamic law, is referenced very frequently by Muslim scholars when providing Q&A services e.g. via online web forums (see section 2.3). Asking questions in English, people use certain Islamic terms in discussion that are not a part of standard contemporary English, e.g. Salat, Wudu, Fard. Such terms are used very frequently and widely understood by most of communicating 'users', e.g. non-Arabic speaking Muslims from the Sub-continent. These terms generally refer to Islamic concepts but these do not have an equivalent translation in English. Hence they are 'imported' into English in their original form. Such words become problematic when processing natural language questions in English for the Quran.

It was also observed that several spellings (*linguistic representations)* exist for imported Islamic terms. This happens because such words are not found in Standard English glosses and dictionaries, and there is no standardised spelling for them. People generate their own transliterated version for these terms. A few of these spellings are used fairly frequently as compared to other alternates; e.g. Saum, Ahram, Mahram, Mahr, Qisas are more frequently

observed spellings for the concepts under consideration. Susan et al note that "A common problem with informative materials related to Islam is incorrect or inconsistent use of terminology..." (Jones 2004). Another study concludes that translators of the Quran have used different words when translating instances of such concepts occurring in Quranic verses. For example the concept of 'As-Sarh' as discussed in verse 027:044 has been translated differently by different authors as follows:

- A glass surface with water underneath it (Khan)
- Lake of Water (Yusuf Ali
- Expanse of Water (Maulana, Sher Ali, Shakir)
- Pool of Water (Pickthal, Rashad, Sarwar)

Existing research involving English translations of the Quran does not address the problem of dealing with imported Islamic terms in English natural language. This problem becomes more complex when the contributing users of the term are neither native Arabic speakers nor they are native English language speakers. For example Muslims living in the Sub-continent do not speak either Arabic or English as a first language. So when asking a question in English they tend to import the local language term used to describe the Arabic concept. This phenomenon can be understood by taking into consideration a Muslim of Pakistani origin. The national language of Pakistan is Urdu. Pakistani Muslims normally refer to the Islamic term 'Salat', an Islamic prayer, as 'Namaz', which is the Urdu word for 'Salat'. Hence, when asking a question in English, this person is likely to import the word 'Namaz' into conversation instead of 'Salat'. In another instance the Islamic term 'Allah' has been spelled differently by speakers of different linguistic origins, e.g. Arabic or Urdu. Further spelling variants for 'Allah' are observed i.e. Allaah and Alaah. There is a trend noticed that users of an Arabic linguistic background tend to spell this word ending with an additional 'h' such as 'Salah' with a final 'h'. On the other hand native Urdu language speakers used entirely different word 'Namaz'.

We believe that above phenomenon exists across all languages and there is a genuine need to fill this gap; *How to deal with imported Islamic terms in English*. We propose a technique: QurTerms as an initial step contributing towards solution. The proposed technique has been applied to an Islamic domain but it could have wider application across domains of similar characteristics to the Islamic domain.

Definitions of ontology insist that an ontology specifically represents common, shared conceptual structures (Brewster and O'Hara 2004). Hence an ontology can be seen as a network of concepts and relationships between these concepts. Thus the technique and devised solution to address the above need have been termed *QurTerms Ontology or*

QurTerms. This Ontology can in effect be considered an Islamic WordNet plus gloss resource. QurTerms holds useful semantic knowledge of three basic types.

- Firstly, QurTerms provides different *Latin-script (English) transcriptions* of the same Islamic terms as used by different authors, translators, scholars and the general public.
- Secondly, QurTerms provides *meanings* for those Islamic terms which do not belong to contemporary English. Hence their meanings are not available in glosses and dictionaries.
- Lastly, QurTerms provides a *thesaurus* for Islamic terms that are commonly used and take different shapes and forms depending on the context of their use. No thesaurus of these words is usually available in the WordNet ontology.

The QurTerms ontology uses XML representation, thus having all the advantages associated with XML representation. The XML schema for QurTerms has been provided in Appendix 2. This resource can be used to resolve imported Islamic terms for natural language processing tasks. It can be adopted for further works that involve such Islamic terms for which similar semantic knowledge does not exist elsewhere. A typical example of a QurTerm is shown in the following figure. A full list of terms included in QurTerms has been made available in the Appendix 3. But it is important to mention that this is not an exhaustive list. QurTerms is the rationale behind the proposed technique which is supported by the QurTerms ontology. Hence it can be adapted to suit the specific needs of a domain application.

```
<QuranicTerm id="42">
  <term> Alamin | Alameen | Raubulalmeen | RabeAlamin </term>
  <meanings>
    <meaning>Lord of mankind, jinns and all that exists</meaning>
    <meaning>Lord of the Universe</meaning>
    <meaning>Lord of the Worlds</meaning>
  </meanings>
</QuranicTerm>
```

Figure 2.28 Example of QurTerm

Section Summary

There are several other computing works involving the Quran which have not been discussed in detail in this chapter but have contributed to the work behind this thesis. Thus include research on XML representation of the Quran in order to make it more accessible and reusable (Taha Zerrouki 2012) and interesting research work published by Aznan et al (Aznan Zuhid Saidin 2005). This work is around a software engineering code of ethics from an Islamic perspective. The research work cited in this section presents how the currently dominant paradigm of scientific tools and techniques in Machine Learning (ML) algorithms, Textual Entailment (TE), Information Retrieval (IR), Knowledge Representation (KR) and Natural Language Processing (NLP) are being applied and tested on the Quranic text as a case study. On the other hand these works also provide evidence of research problems faced due to the peculiar nature of the Quranic text.

Chapter 3 – PARMS

Question Answering (QA) is a shared task for the Information Retrieval (IR), Information Extraction (IE), and Natural Language Processing communities (NLP) (Xu and Klakow 2010). From a general perspective a Question Answering System can be described as an automated process capable of understanding the user question asked in natural language and responding to the question need (D. M. Aliod 2010). The evolution, general architecture and overview of a few recent QA Systems have been discussed in chapter 2 section 2.1.

Traditionally *Answer* is expected to be a factoid piece of information. But several QA systems exist which do not meet this classical definition of *Answer*. Diego et al (Rodrigo, Perez-Iglesias et al. 2010) state that "QA activity relies mainly on a combination of IR and NLP techniques trying to match a natural language query to text snippets...". Most QA Systems embed IR and NLP techniques to achieve QA and this involves identifying the specific paragraphs of the document that contain the information required to answer the question or contain the answer itself (Rodrigo, Perez-Iglesias et al. 2010, Xu and Klakow 2010). This is sometimes called *Paragraph Acquisition*. ResPubliQA 2010 (Peñas, Forner et al. 2010) proposed two tasks related to QA for evaluating the system. Participating systems were required to retrieve a *Paragraph* containing the A*nswer* (Correa, Buscaldi et al. 2010, Glöckner and Pelzer 2010, Pakray, Bhaskar et al. 2010, Pérez-Iglesias, Garrido et al. 2010, Rodrigo, Perez-Iglesias et al. 2010, Sabnani and Majumder 2010). The proposed evaluation methodology states that an answer can be one full paragraph for Paragraph Selection (PS) tasks. One of the aims of ResPubliQA 2010 was set to allow QA technologies to be evaluated against the IR approaches which indicate active research in this area.

We believe that Question-Answering on religious text is likely to be different from traditional Question-Answering on texts collected from television programmes, news wires, conversations, novels and other like sources. Traditional Question-Answering systems differ from QA systems aimed at Question-Answering involving the religious domain in the following few aspects:

- **Nature of anticipated Questions:** Most QA systems aim at answering fact-based questions hence questions are aimed at identifying some factual information. These questions are usually of the interrogatory type, i.e. What, Who, Why, How. On the other hand, it is observed that questions in religious domain are usually knowledge seeking as compared to probing any fact.
- **Nature of Religious Corpus:** Religious text usually exhibit peculiar characteristics that make them challenging but interesting. The Quranic text for example displays figurative use of language. The Quran contains knowledge of rulings around religious concepts and religious viewpoints. Hence the user questions aim to probe certain aspects of a particular concept.

- **Nature of Expected Answer:** Alongside the nature of the question and corpus, the nature of the answer to a religious text is likely to be different also. An answer for a usual QA system is likely to be an extracted fact. But the answer to a religious question (particularly where this is complex in nature) is usually a sub-section (paragraph) of the religious text itself, a Quranic verse for instance. Moreover the user expectation and consumption of this answer is also different. A 'factual' answer is usually consumed by the user *as it is*, whereas the answer to a religious question is firstly *analysed* by the user for its relatedness to their question. Later they *deduce knowledge* from the answer and determine its impact in the context of situation under question.

Due to the above stated differences we believe a traditional QA System requires adaptation for the religious domain. Besides, Monz states that the anticipated user of the Question-Answering System, the type of questions, the type of expected answers and the format in which the available information is stored together determine the design of the Question Answering System (Monz 2003). However, the author believes that so far, no fully-fledged research has been conducted with the aim of developing a methodology to achieve Question-Answering on texts with peculiar characteristics attributed to the Religious domain.

This thesis proposes a *PARMS* (**Par**allel Corpus **M**ulti **S**tream) methodology that attempts to fill this gap. PARMS is a novel method applying existing advanced IR (Information Retrieval) techniques combining them with NLP (Natural Language Processing) methods and additional semantic knowledge to implement QA (Question Answering). A block diagram of PARMS Methodology is shown in figure 3.1. PARMS uses a *Parallel Corpus* and transforms it to yield a merged form as shown in Phase I in figure 3.4. *'All Parallel corpus'* refers to the use of multiple forms of the same corpus. Each form differs from others in a certain aspect, e.g. translations from one language to another by different translators. Additional semantic knowledge can be referred as a *stream* of information related to a corpus, such that PARMS also uses Multiple Streams of this semantic knowledge (ontological form). This additional knowledge can either be used in embedded form to enrich the corpus or in decision making during processing.

The core method of PARMS comprises of phases I to V, as shown in the block diagram, after transforming multiple corpuses into a merged form. Work in this thesis applies the PARMS methodology to a parallel corpus of religious texts. The Quran – the core text of Islam; has been used as a case study (introduced in Chapter 2). Individual verses have been used as paragraphs to be treated as answers. Although components of a QA System depend heavily on the nature of the corpus, the PARMS method has wider application (as discussed in chapter 1). This thesis also implements the PARMS QA Application as proof-of-concept for the PARMS method. The PARMS Methodology has also been evaluated empirically.

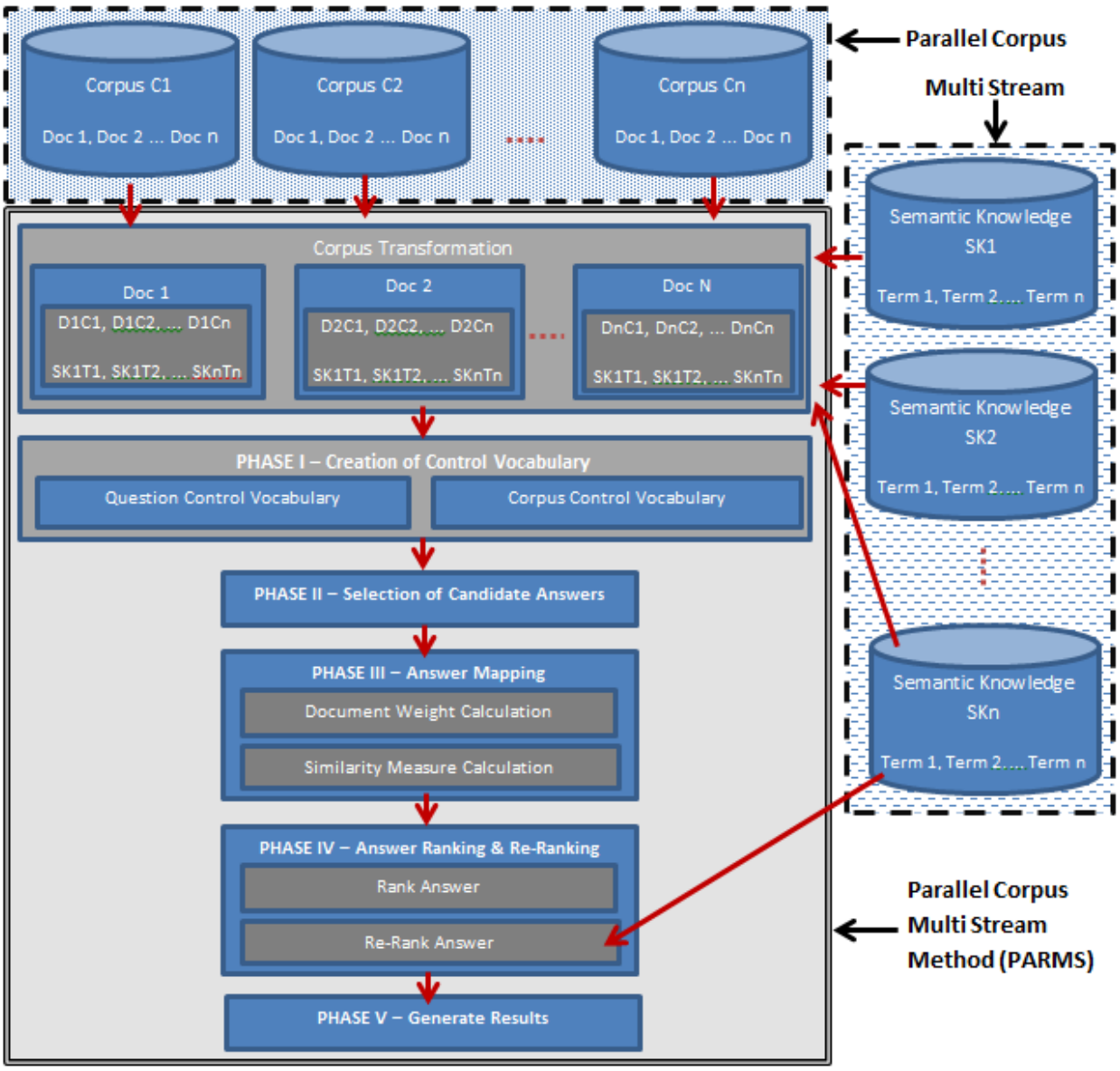


Figure 3.1 Block Diagram depicting Parallel Corpus Multi Stream (PARMS) Methodology (Jilani 2013)

Details of each of parallel corpus processing, semantic knowledge enrichment and phases that constitute the PARMS method are provided later in this chapter. The theoretical aspects of IR and NLP that the PARMS Methodology uses have been introduced briefly at the beginning of the chapter. The next section provides details of each phase of the PARMS methodology. Implementation details of the PARMS QA Application are provided towards the end of the chapter. The next two chapters of the thesis provide details of the evaluation of the PARMS methodology.

3.1 Theoretical Underpinnings

This section describes different concepts and methods of Information Retrieval (IR) and Natural Language Processing (NLP) investigated and exploited by the PARMS methodology. PARMS applies a unique combination of these concepts and methods, to map a user

question, expressed in natural language, to (a) verse(s) in a Quran translation that is/are most relevant as an answer.

3.1.1 Control Vocabulary – The Metadata

Control Vocabulary (CV) is the metadata about the corpus (Quran translation) or the user query (the question). The purpose of producing CV is to provide a foundation for mapping a user query to the most relevant verse(s) in the Quran translation.

Control Vocabulary comprises "bag-of-words" *derived* from the documents (corpus or the user query) or added as part of term expansion, section 3.1.5. Not all the words in these documents are eligible to be included in the control vocabulary. Instead there are certain words that are disregarded during CV generation. For example stop words, section 3.1.3, are considered less significant, hence are ignored in this process. Also the words to be included are transformed to their root form, section 3.1.4, before adding them to the control vocabulary. Each word in the control vocabulary is called a ***term***. The Control Vocabulary records key statistical information about each term. This information includes the following:

- List of documents in which the term exists
- Frequency of the term in each document
- Size of the document in which the term exists
- Weight of the term in each document

Different weighting functions can be applied to calculate the weight of a term within each document. These weighting functions are described in the next section.

3.1.2 Weighting Functions

The PARMS QA system is capable of applying different weighting functions to quantify the importance of a term within a document. This quantitative value is termed *weight* and is indicative of the importance of a particular *term* within a document as compared to other terms within the same document. Each *term* within the control vocabulary is attributed by a *weight* that is later used to measure the degree of relatedness between a verse and the user query.

(i) Term Frequency – (tf)

Term frequency (Jones 1972, Salton, Fox et al. 1983, Salton and McGill 1983, Salton and Buckley 1988), denoted by $tf_d(t)$, is a simple measure of raw frequency of a term in a document. It is simply the number of times that term ***t*** appears in document ***d***. In other words it is the frequency-count of the term ***t*** in the document ***d***.

$$\mathrm{tf_d(t)} = \mathrm{f_d(t)}$$

Where $\mathbf{f_{t,d}}$ is the frequency of a term ***t*** in a document ***d***.

(ii) Inverse Document Frequency - idf

Inverse Document Frequency (Jones 1972, Salton, Fox et al. 1983, Salton and McGill 1983) is a measure to signify the importance of term within all documents (Svenonius 1972) (Jones 2004) (Noordin and Othman 2006) (Yusof, Zainuddin et al. 2009). The Inverse document frequency of a term ***t*** in a corpus **C** can be calculated as:

$$\mathrm{idf_d(t)} = \log \frac{\|C\|}{\mathrm{df_t}}$$

Where $\|C\|$ is the total number of documents in the corpus **C** and $\mathbf{df_t}$ is the number of documents in which term **t** exists.

(iii) Term Frequency – Inverse Document Frequency

Term Frequency – Inverse Document Frequency (Jones 1972, Salton, Fox et al. 1983, Salton and McGill 1983, Salton and Buckley 1988, Wu, Luk et al. 2008), denoted by tf-idf, is a simple product of term frequency and inverse document frequency calculated as follows:

$$\mathrm{tf-idf_d(t)} = \mathrm{tf_d(t)} \times \mathrm{idf_d(t)}$$

(iv) OKAPI Model

The OKAPI (Robertson, Walker et al. 1993, Beaulieu, Gatford et al. 1997, Robertson and Walker 1999, Robertson, Walker et al. 1999) weight model calculates the weight of a term in the document and user question using the following formula respectively (Di Cai and Lee McCluskey 2011).

$$W_d(t) = \begin{cases} \dfrac{(a+1)f_d(t)}{a \times \left[(1-b) + b\,\dfrac{\|d\|}{avc(C)}\right] + f_d(t)} & when\ t \in V_d \cap V \\ 0 & when\ t \in V - V_d \end{cases}$$

$$W_q(t) = \begin{cases} \dfrac{(c+1)f_q(t)}{c + f_q(t)} \times \log \dfrac{|C| - [F_C(t) - d]}{F_C(t) - d} & when\ t \in V_q \cap V \\ 0 & when\ t \in V - V_q \end{cases}$$

Where ***a*** = 1.2, ***b*** = 0.75, ***c*** = 1000 and ***d*** = 0.5.

$\boldsymbol{f_d(t)}$ is the frequency of a term in the document

$f_q(t)$ is the frequency of a term in the question
$F_c(t)$ is the frequency of a term in the corpus
$avc(C)$ is the average size of corpus **C**
C is the size of the corpus

3.1.3 Stop words

Stop words are those words that are considered less significant and, therefore, need to be filtered out at some stage during processing. These words are insignificant either in the context of a language or due to the domain in which they are being used. Examples of language specific stop words include short function words, such as 'the', 'is', 'at', 'which', and 'on'. Some of the words within a corpus may be so frequent that they do not play a significant role in classification. Such words are examples of domain-specific stop words. During processing, the presence of these words not only consumes valuable processing resources but may also distort the results by skewing the statistics in frequency-based methods. There is no definite list of stop words as different systems use different stop words according to their requirements. The PARMS Methodology has been tested using a list of stop words compiled by Stanford University with a few additions, e.g. the word 'Allah'. PARMS removes these stop words early on in the process to minimise their effect.

3.1.4 Stemming

An important part of the Question Answering (QA) system is the Information Retrieval (IR) task that requires determining which document in a collection of documents contains the required information. Usually basic information retrieval tasks involve a keyword based search query. A document could contain a few keywords or all the keywords in the query. Words in natural language occur in different morphological forms. For grammatical reasons most words in English are derived from their respective root. This phenomenon is also reflected in the written text. Hence the words in the documents are mostly variants of the basic root words, e.g. *"computer"* and *"computes"* are both derivations of the root word *"compute"*. There are other situations where words belong to the same family and all lead to the same meaning e.g. *"democracy", "democratic"* and *"democratisation"* are related words, so a keyword search for *"democracy"* should also pull up documents which contain the other related words.

To achieve the above, words need to be reduced to their root (non-inflected) form. In cases of related words they need to be reduced to the base form of the word by *Stemming*. Stemming is the process of normalising words in the text to their stem. The aim of using Stemming is to improve the IR. It is not a grammatical requirement. In fact stemming can lead to mistakes. Stemming often involves removing the 'endings' or suffixes of words

bringing them to their root form. Stemming does not apply similarly to all languages, e.g. Chinese and Arabic. A programmatic implementation of stemming is termed a *stemmer*. There are different stemmers currently being used. PARMS is capable of using the following stemmers.

- SnowBall Stemmer
- Porter Stemmer
- Pling Stemmer

3.1.5 Term Expansion

Term expansion is the process of reformulating the seed query in order to improve the performance of the information retrieval task. This is generally done by adding additional keywords that are related to the original query. Different techniques can be applied to achieve this.

Query Expansion

In Query Expansion additional keywords are added to the original user question. A common technique is to use synonyms of the keywords in the original question. An example of this technique is to use WordNet (Miller, Beckwith et al. 1990, Miller 1995, Fellbaum 1998, Pedersen, Patwardhan et al. 2004) to find out synonyms of a keyword used in the question and then add or replace these synonyms in the original question. Another way could be to use any additional semantic knowledge related to the question's keywords and incorporate that in the original question. An example of such a technique is to use domain-specific ontologies that may contain more specific information about that keyword.

Corpus Enrichment

Query enhancement can be applied to the corpus documents. Domain specific ontologies can be best suited to achieve this. Anaphoric resolution can be seen as a technique that potentially enriches a corpus document. The PARMS Methodology takes advantage of this assumption and uses domain specific ontologies for corpus enrichment.

3.1.6 Similarity Measure

Similarity measure is a quantitative value that represents the degree of relatedness between two documents. In the context of a QA system the user question and the corpus document under consideration can be assumed to be the two documents that need to be compared. This measure is generally a normalized value between 0 (zero) and 1 (one) where 1 (one) represents the highest level of resemblance and 0 (zero) indicates irrelevant documents. In a rank-based QA system, the similarity measure is calculated for every document in the corpus. As a next step, these documents are ranked using their relatedness

to the user question. The document rank is then used to decide its eligibility for further processing and inclusion in the final result. There is a variety of ways to calculate the similarity measure and their choice depends on the nature of problem being solved. PARMS is capable of using following two similarity measures:

Cosine Similarity

Cosine similarity (Singhal 2001, Manning, Raghavan et al. 2008) measure is used to measure the similarity between two vectors. Representing user question and corpus document as vectors, cosine similarity can be measured as:

$$Cosine = \frac{\sum_{t \in V}(W_d(t) \times W_q(t))}{\sqrt{\sum_{t \in V} W_d^2(t)} \times \sqrt{\sum_{t \in V} W_q^2(t)}}$$

Where $W_d(t)$ and $W_q(t)$ is the weight of term t in document d and question q respectively.

LESK

The LESK algorithm (Lesk 1986) is basically used for word sense disambiguation. LESK Algorithm is based on the assumption that the words within a "neighbourhood" are most likely to share a common concept. So a count of common words between a specific *sense of a word* and its neighbouring words can help disambiguate the sense of that particular word from all its existing senses. The *sense of a word* having a higher count of words in common is likely to represent the true sense of that particular word. This rationale for the LESK algorithm can be adopted and applied to other similar problems. For instance, the same principles can be applied to measure the relatedness between the user question and the corpus document. This would mean that document(s) having a higher number of terms in common with the user question are likely to be related. PARAMS adopts the LESK algorithm to help improve the results such that it uses a measure of term weights instead of mere word frequencies to measure relatedness.

3.2 Methodology

PARMS aims to map a user query to the most relevant verse(s) as an answer from the Quranic corpus. This objective is achieved by applying a series of operations to the *documents* to find ranked answers. Here the term *document* may refer to either 'a verse' or 'a user question'. Standard Information Retrieval and Natural Language Processing techniques and resources have been used to devise these domain-independent operations. Additional, domain-specific, semantic knowledge resources have been utilised to improve the results. These operations can be classified as distinct *phases* for better understanding. Each of these phases performs a specific operation to produce an output which is used as

the input for the successive phase. The PARMS Methodology comprises the following five phases:

1. Creation of Control Vocabulary
2. Selection of Candidate Answers
3. Document Weighting & Similarity Measure
4. Answer Ranking & Re-ranking
5. Generation of Results

The PARMS phases listed above also require some pre-processing to transform parallel corpuses into a merged form. During this processing all corpuses are read and their corresponding documents are appended together. Each of these merged documents is given the same reference as its original document in the corpus. This transformation process has been described with an example in Section 6.2 under Parallel-Corpus Mode. These merged documents are used in further processing to produce the answers. The reference numbers of these answers are used to cross-reference the original document from any of the corpuses to fetch the original document for the answer. The PARMS phases are described as under the following headings:

3.2.1 Phase I – Creation of Control Vocabulary

The creation of Control Vocabulary (CV) is the key phase that provides a foundation for later phases. Although the CV generation process is the same for CV creation for both *documents,* i.e. the corpus and the user question, a separate CV is created for both. This is because the two vocabularies use different semantic knowledge resources. The creation of control vocabulary is carried out in the following three steps:

Step - 1

The first step is performed on each *document,* i.e. the corpus and the user question. The document contents are tokenized by chopping on whitespaces and the punctuation characters are discarded. *Term expansion* is also performed at this stage. During *term expansion*, semantic knowledge related to the document is also tokenized. Care is sought while tokenizing in order to avoid duplicates in these extended tokens. This is because duplicates may distort the statistics at a later stage.

As mentioned previously, CV creation of PARMS uses two semantic knowledge resources to enrich the user question. These are the *WordNet* (Miller, Beckwith et al. 1990, Miller 1995, Fellbaum 1998, Pedersen, Patwardhan et al. 2004) and *QurTerms* ontologies. Each token, produced from an original question, is queried against *QurTerms Ontology* and its translation, if it exists, is picked up and tokenized. Similarly each token is also fed into the

WordNet Ontology and resulting synonyms, if any, are also added in as tokens. As *WordNet* requires a POS tag for each word being queried, so POS tagging is also performed on each token before passing it on to *WordNet*.

Semantic knowledge added to the corpus documents is borrowed from *QurAna Ontology* and optionally from *WordNet*. For each document (Quranic verse) *QurAna* is queried, using a document reference number, to find out if there are any pronouns related to the particular document. If these are found, the pronouns are also tokenized and added to the list of *tokens* of that particular document (Quranic verse). This final list of *tokens* is termed a control vocabulary document, *CVDoc*. Each *CVDoc* is given a unique document reference number. The final outcome of this step is a list of *CVDoc*s, indexed by document reference, containing its related *tokens*.

Step - 2

The second step starts by converting *tokens* into *terms*. This is done by traversing the list of tokens for each *CVDoc* and removing the tokens deemed to be *stopwords*. The remaining tokens are then normalised and stemmed to produce the *terms* related to each *CVDoc*. There is one-to-one mapping between *CVDoc* and the original document. In fact, *CVDoc* is just another formal representation of the same document with some additional statistics attached to it.

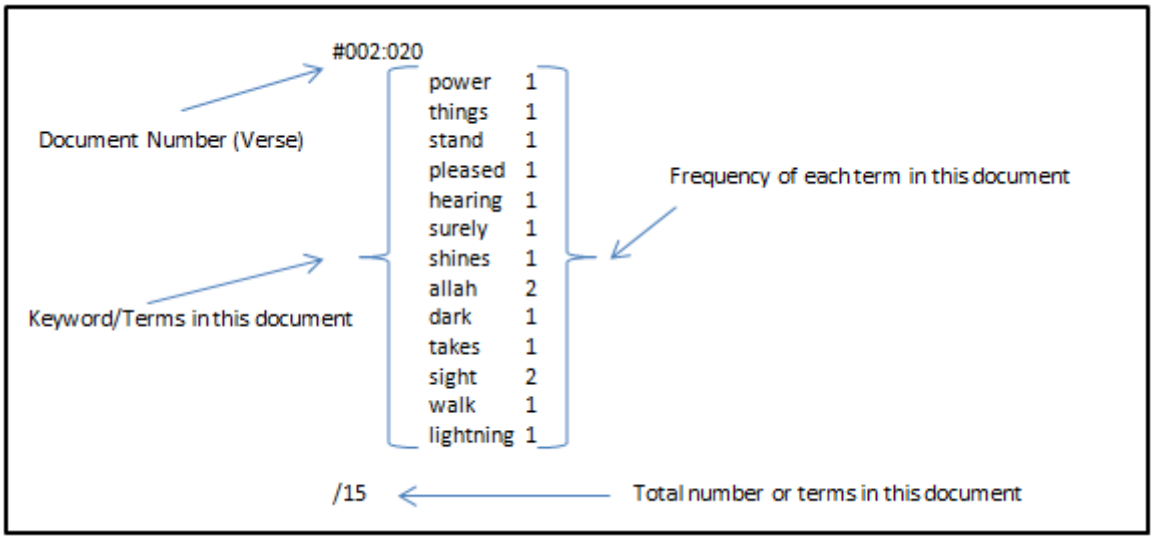


Figure 3.2: Level 2 Control Vocabulary (Jilani 2013)

It is important to note that *CVDoc* acts as a proxy of the original document during processing. Hence it is valid to use the terms 'document' and '*CVDoc'* interchangeably in a processing context. Further to this the resultant *terms* are aggregated within each *CVDoc*. Some key statistics are calculated for each *CVDoc* in the list. These statistics include the total number of terms (i.e. size of the document) and frequency of each term within this

CVDoc. The final outcome of this step is a list of *CVDoc*s, indexed by document reference as key, where each document contains its related terms and statistics. This list is called the *Level 2 Control Vocabulary* (Appendix 5) as shown in above figure 3.2 as an example.

Step - 3

The third and final step is used to convert control vocabulary from *Level 2* to *Level 3*. This is done by collating each term across all *CVDoc*s and recording the document references in which the term exists. This structure contains the *term* and its related document references in *term* as Control Vocabulary Term, *CVTerm*. Like *CVDoc*, *CVTerm* also acts as proxy of a *term;* thus these can be used interchangeably as well. *Level 3* is effectively the transposed form of *Level 2,* such that Level 2 contains a list of documents and their related terms whereas Level 3 is a list of terms and their related documents.

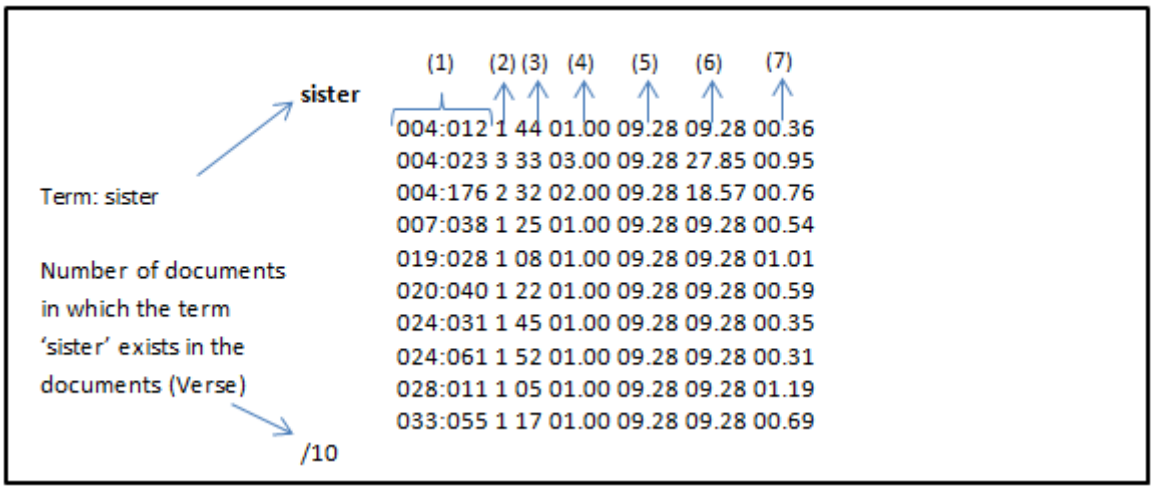


Figure 3.3: Level 3 Control Vocabulary Example - List of documents the term 'sister' exists in and weights (Jilani 2013)

The data shown in figure 3.3 depicts following information:

(1) Document ID
(2) Number of occurrences of term ('sister') in document d
(3) Number of terms in document d
(4) Term Frequency of keyword for use in TF/IDF score
(5) Inverse Document Frequency for use in TF/IDF score
(6) TF/IDF score taking account of 4 and 5
(7) *Okapi score, which takes account of 3, 4, and 5*

At this point, more statistical measures are also calculated and attributed to each *CVTerm*. These measures include the number of documents in which the term exists and most importantly the *term* weight. *Term* weighting is carried out using a selected weighting function to assign weight to each term within each document. The final outcome of this step is the list of *CVTerms,* indexed by the *term* as a key. Each *CVTerm* has its related document

references and statistics. Figure 3.3 shows the example of *Level 3 Control Vocabulary* (Appendix 5). Figure 3.4 provides a block diagram of the Control Vocabulary (CV) generation for Questions as well as the corpus. The semantic resources used by question and corpus differ. Hence they have been shown separately until the token creation process.

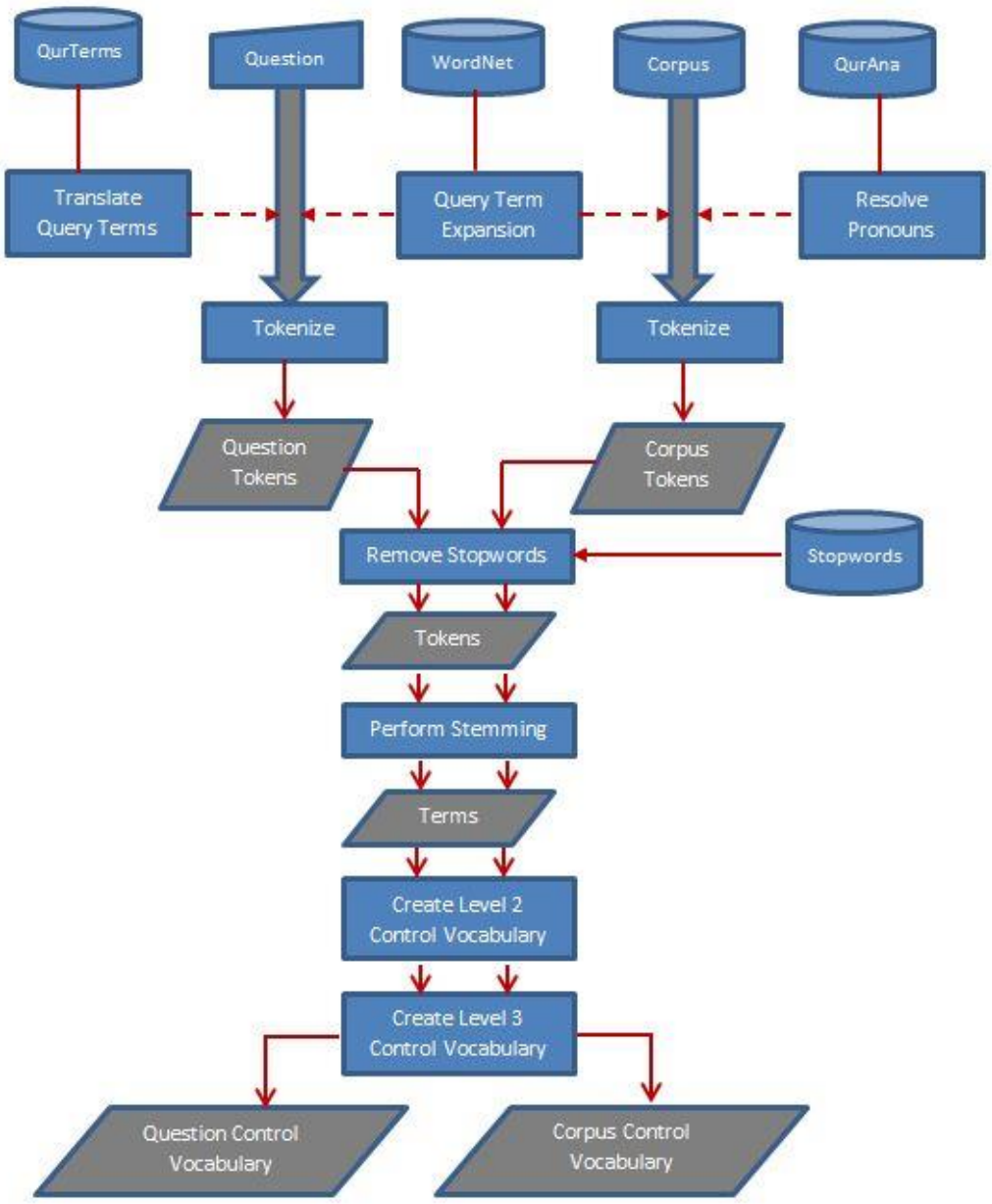


Figure 3.4: Block Diagram depicting Control Vocabulary Generation (Jilani 2013)

3.2.2 Phase II – Selection of Candidate Answers

The PARMS Methodology uses this phase to extract any documents from the entire corpus that potentially have any relevance to the user question. These extracted documents are called *candidate answers*. The objective of this phase is to identify if there is any document

(verse) that is relevant in the context of the user question. If there is, include this document in the list of candidate answer pool to compete for the selection as final answer. The degree of relevance of the document does not matter; even documents with a small level of relevance are included. As a result irrelevant documents are filtered out at this early stage whereas all relevant document(s) are included in the candidate answer pool. These candidate answers are used for further processing.

This phase requires *Level 3 Control Vocabulary* for both corpus and user question. Using both of these control vocabularies, the user question is mapped to all the relevant documents. This is done by comparing all *CVTerms* of the user question, one by one, with the corpus's *CVTerms.* If any of the corpus's *CVTerms* is matched then all its related documents are added to the candidate answers pool.

It is important to note that at this stage there is no measure of degree of relatedness for the candidate answers. All the candidate answers are equally likely to be selected as the final answer. On the other hand, all irrelevant documents have been eliminated from the selection pool. This will help focus further processing only on those documents that stand any chance of selection as the final answer.

3.2.3 Phase III – Document Weighting & Similarity Measure

This phase works in two successive steps. Firstly, each document in the candidate answers pool is assigned a weight using *Document Weighting.* Secondly, the similarity measure for each pair is calculated using a suitable method.

Step – 1

In document weighting, the candidate answers extracted in the previous phase are collated in such a way that the weights of all the terms of a document that had matched with any of question's terms are added to document's weight. In other words, a document's weight is the cumulative weight of its terms that are common to the document and the question.

Step - 2

The similarity measure is the degree of relatedness of a candidate answer to the user question. In this step this measure is quantified for each of the candidate answers. In the LESK based similarity measure, raw document weight can be treated as a similarity measure. It is based on the premise that the higher the number of common terms between the question and a candidate answer the higher the similarity measure. This idea is further enhanced by using *term* weights instead of a *term* count. So a document with *term(s)*

having higher weights is more related to the question than a document with a similar number of *terms* but lower *term* weights. In the CoSine similarity method, *term* weight metrics of both the question and each candidate answers are used to calculate the similarity measure.

3.2.4 Phase IV – Answer Ranking & Re-Ranking

In this phase candidate answers are ranked according to their similarity measure determined in the previous phase. The higher the similarity measure, the higher will be the rank and vice versa. Candidate answers are sorted in descending order of their similarity measure and are assigned a rank equivalent to their position in the list.

Re-ranking in PARMS is aimed at improving the ranks of candidate answers in order to increase the probability of finding a correct result. This is done by boosting their weight using additional semantic knowledge, usually domain specific. As the PARMS Methodology has been applied to the Quran as a case study, the domain-specific ontology has been used for the purpose, i.e. the *QurSim Ontology* (see section 2.3). QurSim provides a list of all similar verses to a given verse based on their semantic similarity. This knowledge can be used to re-weight the candidate answers found at higher ranks by borrowing the weight of 'lower ranked' candidate answers that are similar to a higher ranked candidate answer. QurSim has different levels of similarity between verses and PARMS is currently using similarity level 2 to ascertain that two candidate answers are similar to each other.

The idea is based on the assumption that as similar answers refer to same concept being queried so the lower ranked candidate answers that stand less chance of getting into the final result can lend their weight to their higher ranked 'siblings' to improve their chances of being selected in the final result. The candidate answers that lend their weight are, obviously, surrendering their 'candidature' and are eventually eliminated from the candidate answer list. In PARMS, re-ranking is performed on a list of candidate answers. These answers are already ranked in descending order of their weight. PARMS traverses this list from top to bottom, selecting an answer one by one, as a '*pivot*', given that the answer has not already lent its weight to another answer. This pivot answer is compared with the rest of the answers, one-by-one, starting from next lower answer down to the end of the list. The weight of the answers being compared with the pivot answer will be added to the weight of the pivot answer if:

- The answer is similar (according to QurSim) to the pivot answer
- The answer has not already lent its weight to another (pivot) answer

If an answer lends its weight to a higher rank sibling this means that the answer is conceding its candidature to the higher ranked answer. As this answer is no longer a contestant for selection so its weight is nullified. This is to ensure that the answer:

- contributes to only one higher ranked sibling
- is not a contender for selection any more.

Once all the answers are processed, as pivot, and re-weighted (if re-weighting is due) then the list is re-ordered in descending order of their weight thus assigning each answer a new rank. The answers that have lost their weight in the process will be pushed down to the bottom of the list and are later excluded from selection. A pseudo-code describing the PARMS re-ranking method has also been presented in figure 3.5.

```
Module ReRank (CandidateAnswerCollection)  Comment: Collection of candidate answers
    Set TotalCandidateAnswers = CandidateAnswerCollection.Size
    For Rank = 1 To TotalCandidateAnswers
        Set CurrentAnswer = CandidateAnswerCollection[Rank]
        If CurrentAnswer.Weight is greater than 0 Then
            Set SimilarVerses = { Get similar verses for current answer }
            For LenderRank = Rank + 1 To TotalCandidateAnswers
                Comment: Start processing from answer next to current answer
                Set PotentialLenderAnswer = CandidateAnswerCollection[LenderRank]
                If PotentialLenderAnswer.Weight is greater than 0 Then
                    Comment: This lower ranked candidate answer has not lent its weight
                    Set Ref = PotentialLenderAnswer.Ref
                    If Ref exists in SimilarVerses Then
                        Comment: This lower ranked candidate answer is similar to current answer so it is to lend its
                        weight to current answer
                        Set PotentialLenderAnswer.Weight = 0
                        Add PotentialLenderAnswer.Weight To CurrentAnswer.Weight
                    End If
                End If
            End For
        End If
    End For
    Reorder CandidateAnswerCollection
End Module ReRank
```

Figure 3.5: Pseudo-code for PARMS Re-Ranking Method (Jilani 2013)

3.2.5 Phase V – Generation of Results

In this final phase, ranked candidate answers are used to pick up the *result*. A PARMS *Result* is the top *N* candidate answers, which now are termed *answers*. These answers are ranked according to their similarity measure score. The answer most related to the question is at the top. Candidate answers that did not make to the top N list are, hence, discarded. The answers here are the *CVDocs* containing reference to the actual corpus documents (verses). Their references are then used to fetch actual Quran verses along with their Arabic counterparts to complete the result.

3.3 Implementation

This section discusses specific requirements, design and implementation of a system as a proof of concept of the PARMS Methodology described above. Christof Monz believes that the anticipated user of the QA System, the type of questions, the type of expected answers and the format in which available information is stored together determine the design of the Question Answering System (Monz 2003). Hence before implementing a system, it is necessary to evaluate the requirements of a software application that can best implement the proposed method. Once the requirements were identified, prospects of using an existing open source framework were also explored.

Apart from general requirements of a QA system, the proposed PARMS's methodology requires some specific features. Both the PARMS question and its potential answer are different as compared to question and answer in a traditional QA system. A question in a traditional QA system is generally a short natural language text that belongs to a specific question type, e.g. 'what', 'where', 'who', etc. On the other hand a question in PARMS is more like a complex question which is a section of text seeking knowledge about one or more specific concepts. Similarly the PARMS answer is a collection of extracts from a religious text ranked by the degree of relevance to the concept being sought by the user query. This is different to a factoid or list answer provided by a traditional QA system.

Flexibility and configurability was another significant requirement from the system implementing the PARMS Methodology. This is because the effect of some of the concepts and features used by the PARMS Methodology had to be explored to find the best performing set of features. Similarly some concepts have multiple implementations available, e.g. stemmer, similarity measure, etc. Hence it was necessary to explore which of these implementations performs better in conjunction with other concepts and features for the specific corpus used by PARMS, i.e. the Quran corpus.

In order to implement a system capable of satisfying the above requirements, OpenEphyra (Schlaefer 2007) was explored as a potential candidate. OpenEphyra is an open-source extendable framework to implement a QA system. Although OpenEphyra is a good framework that can be extended to fulfil specific requirements, it is designed to implement a traditional QA system. Exploring the possibility of implementing PARMS within OpenEphyra, it was found that this could require disabling or even dismantling the core features and components of the OpenEphyra framework. Doing so would defeat the basic purpose of using an existing framework. So a bespoke system was required to execute the proposed Methodology suitable for the requirements discussed above. The PARMS Question Answering System was implemented as proof of this concept.

PARMS QA system has been developed as a configurable desktop application. It can be configured to select a particular implementation of certain concepts and features like term weighting function, similarity measure method, etc. Similarly different operations like removal of stopwords, perform stemming, etc. can also be enabled or otherwise using configuration options. Choice of corpus is also configurable whereby a particular or all translations of the Quran can be used as corpus by the system to test PARMS Methodology.

3.3.1 System Design & Development

The PARMS QA system is designed and developed using standard software engineering practices. An evaluation framework is also developed for initial experimentation to explore the effect and implementation of different concepts. The system is developed using Eclipse IDE and Java programming languages. The PARMS QA system has been implemented using three distinct layers. Each layer is responsible for a specific task. This "separation of concerns" is used to make sure that each layer should work independently of every other layer and can be changed, if required, without affecting the other components of the system. Details of these layers are given below:

Data Layer

This part of the system is responsible for storage and access of data. Data includes the corpus (Quran translations), ontologies (QurSim, QurAna & QurTerms) and test questions (used in evaluation). This data along with the system configuration is stored as XML. **J**ava **A**rchitecture for **X**ML **B**inding (JAXB - *https://jaxb.java.net*) is used to bind and read (*unmarshal*) these data files into Java objects for further use.

Processing Logic Layer

Implementing the PARMS methodology, this layer constitutes the core of PARMS QA system. All phases of the PARMS methodology, from creation of control vocabulary to generation of results are performed by this layer. This layer uses data and configuration provided by the data layer to execute the methodology steps.

Presentation Layer

This layer provides UI for system interaction. The user question can be entered using this layer that passes this on to the processing layer. After querying the results are passed back to the presentation layer. These results are then displayed using standard windows UI controls. In the evaluation system this layer is also used to select different configurable options. The graphical User Interface (GUI) provided by this layer is made up of Java *Swing* components and is designed with the help of the WindowBuilder plug-in.

3.3.2 PARMS Question Answering Application

The PARMS Question Answering Application implements the PARMS methodology. It allows the user to perform a Quran-related query to find potential answer(s). The application takes in a user question in natural language (English) and provides top *N* relevant verses as answers.

The application uses a pre-set configuration to select different parameters for particular concepts and methods used in the PARMS methodology. On start-up, the application executes Phase I of the PARMS methodology, as shown in figure 3.1 & 3.6 (details provided in section 3.2.1) for the corpus using selected parameters. On completion of this phase, the user is presented with a UI and the application is ready to take and process user questions.

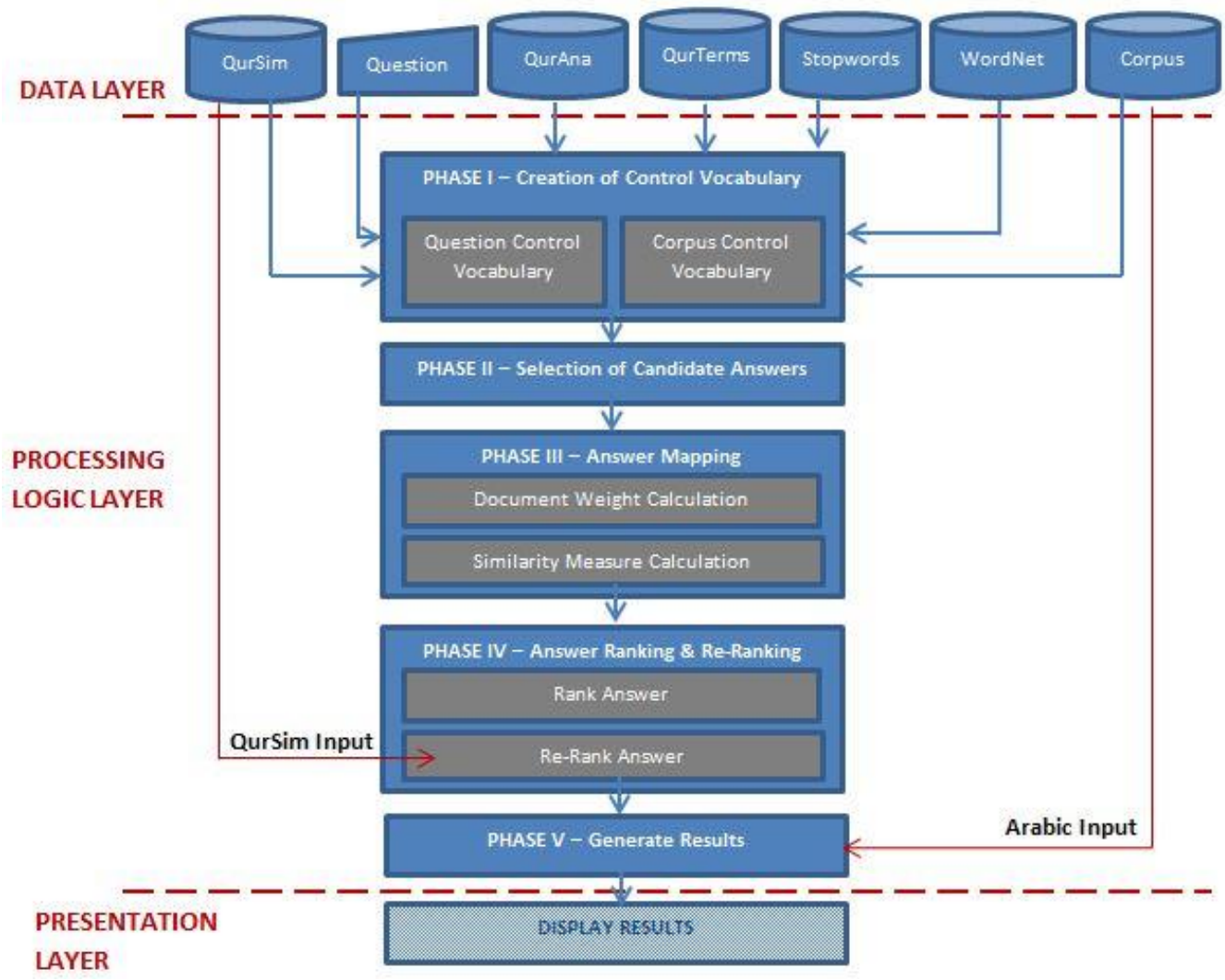


Figure 3.6: Block Diagram depicting PARMS QA Application Architecture (Jilani 2013)

When a user question is entered, the application performs Phase I of the PARMS methodology for this question. This is followed by execution of phases II to V of the PARMS

methodology to generate the answers. At the end of this process, the resultant answers are presented to the user. The architecture of the PARMS QA Application has been presented in figure 3.6. This diagram also depicts the three different development layers i.e. the Data Layer, Processing Logic Layer and Presentation Layer. It also indicates the PARMS Methodology Phases I – V. Details of the individual phases have been provided in the previous section.

The screenshot presented in figure 3.7 shows the Graphical User Interface (GUI) for the application. The answers are displayed using the lower panel in the above screen shot. The result panel displays answers both in English and Arabic versions of the answer verses. Each English answer row contains a title. This title comprises the chapter name, the Arabic name of that chapter in transliteration, the English name of that chapter and finally the verse number. The title for the Arabic answer displays the chapter name in Arabic and the verse number.

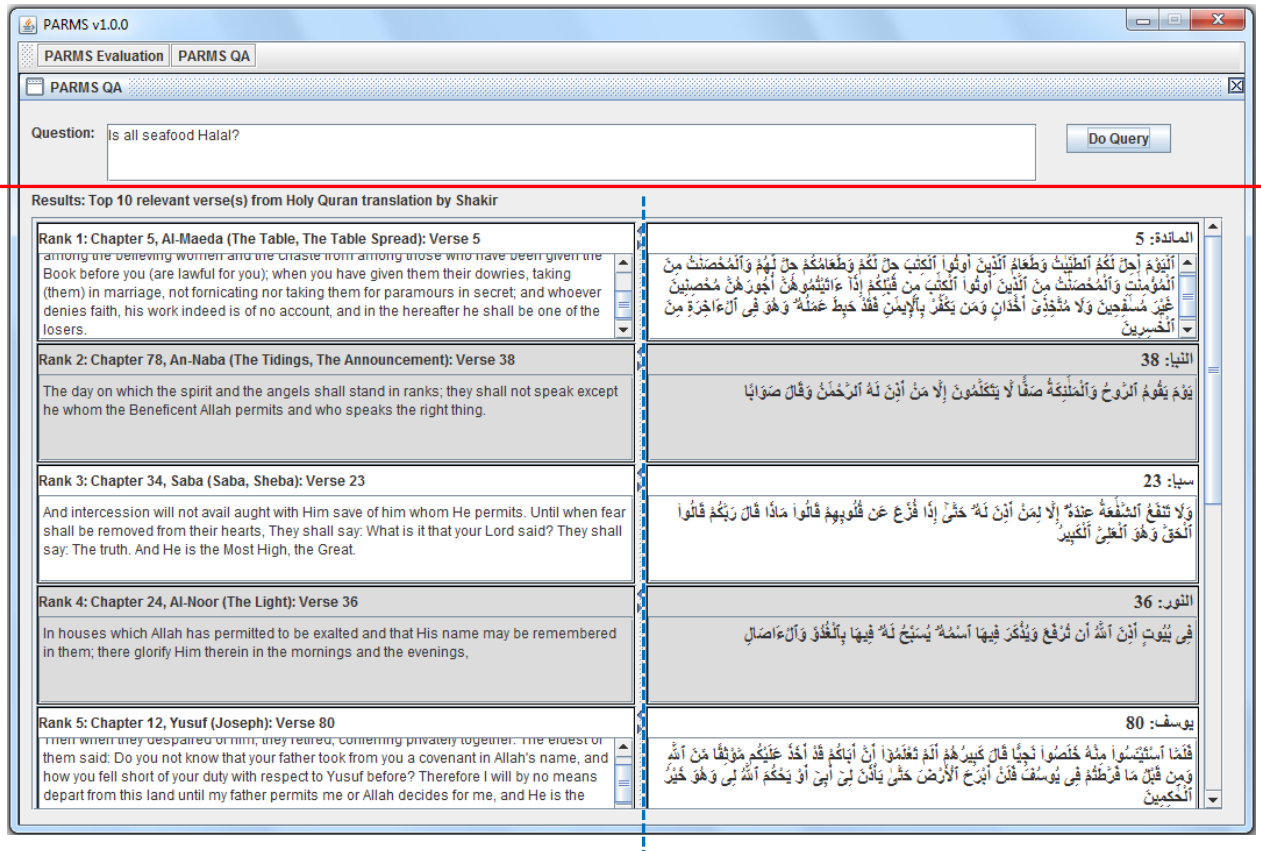


Figure 3.7: Screenshot of PARMS Question Answering Application (Jilani 2013)

3.3.3 PARMS Evaluation Framework

The PARMS Evaluation framework is used to experiment with different parameter settings for different concepts and methods utilised by the PARMS methodology. It helps to study

the effect of different concepts and methods by selecting their different variations or by switching them on/off. Details of the parameters related to different concepts and methods and available options have been provided in Chapter 4.

The PARMS evaluation framework is a batch processing application. It can process a batch of pre-defined natural language questions related to the Quran. The framework finds the answers to each question using the PARMS methodology and then evaluates the batch of results by comparing the resultant answers against pre-defined *correct answers* for each question. Different statistics are also produced by this framework to quantify the analytical findings. Evaluation metrics and results of experiments are discussed in detail in the next two chapters respectively. Evaluation results are also logged in a spread sheet and different text files for further analysis. A sample of these are available in Appendix 4.

On start-up, the application presents the interface as shown in the screenshot below. The panel on the left-hand side is used to select different parameters used within the PARMS methodology. After parameter selection, Phase I of the PARMS methodology is executed to produce control vocabulary for the corpus. Afterwards, questions (PARMS Model Question Set) are read from a disk file and processed one by one.

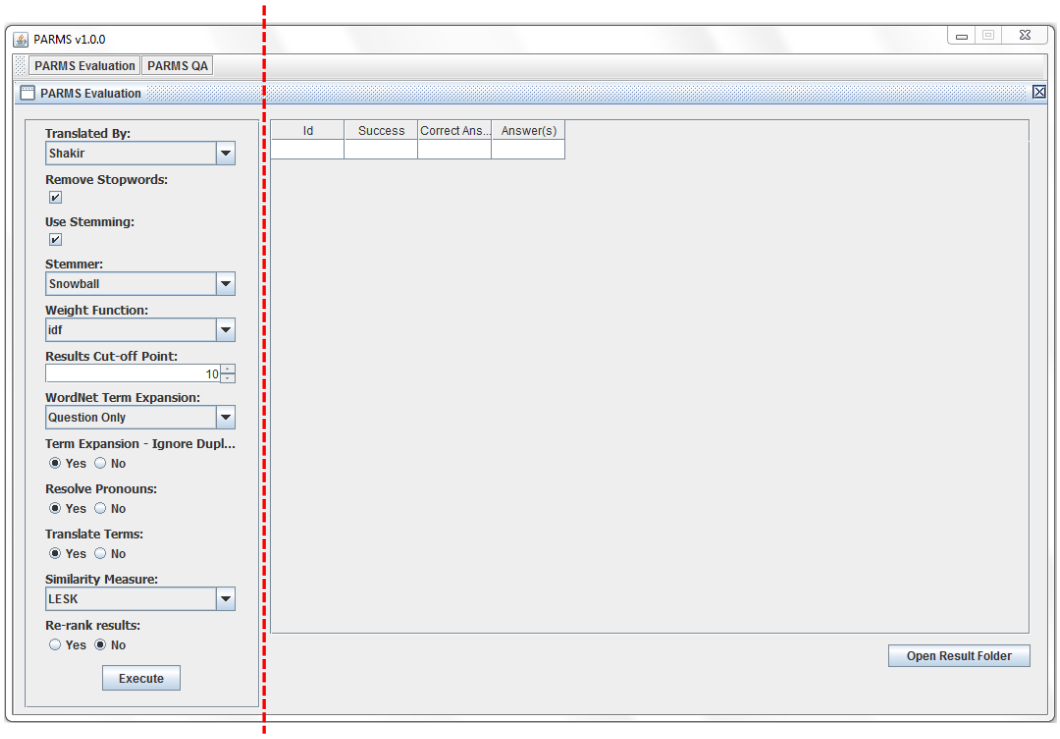


Figure 3.8: Screenshot of PARMS Question Answering Evaluation Framework (Jilani 2013)

Processing of each question involves creation of its control vocabulary (PARMS methodology phase I) followed by execution of the rest of the Phases II - V using corpus' control vocabulary. The resultant answers for each question are then used to collate the results of

the whole batch. The batch results are then evaluated and quantified using different statistical measures. A log of these collated results is also saved to disk in spread sheet format for any further analysis. A sample of this file is available in Appendix 4.

The right-hand side panel is used to display the summary of evaluation results. The evaluation result summary of each question is represented by a row in the evaluation results table. The following screenshot displays the results of experiment.

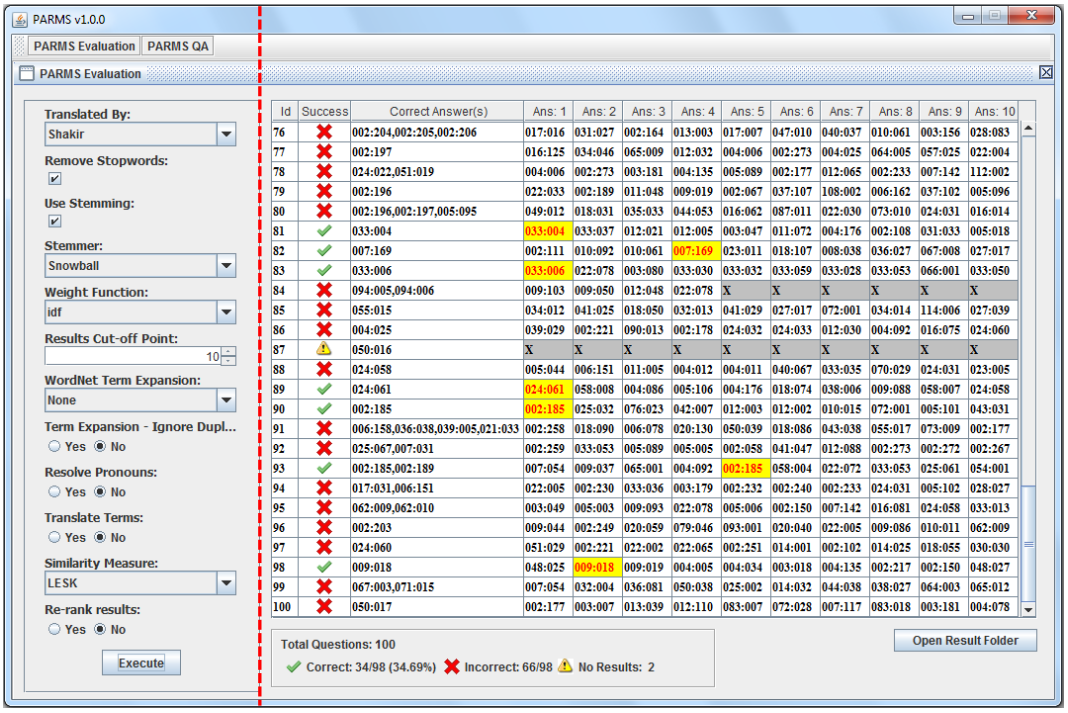


Figure 3.9: Screenshot of Evaluation Results using PARMS Model Questions Set (Jilani 2013)

The details of each entry in the evaluation result summary are as follows:

- Id – question ID number within question set
- Success – icons indicating correct, incorrect or no answer(s)
- Correct Answer(s) – One or more pre-identified correct answers separated by comma (,)
- Ans: *n* – top *N* ranked answers from 1 to *N*. A correct answer is highlighted while a cross (**X**) with grey background indicates that no answer was found.

3.4 Testing

Testing is an investigation conducted to verify quality and performance of a piece of software. General testing strategy is usually to compare the results of a program against known values, but this was not suitable for testing PARMS evaluation tool. This is due to the nature of the system. Hence a testing strategy was devised for the purpose. The output of

PARMS evaluation tool was compared against known values that had been generated by *manual* processing of question and potential answers through PARMS methodology. Parameter settings used for testing are as follows:

Translated By	Remove Stopwords	Stem Words	Stemmer
Shakir	Yes	Yes	Snowball
Weight Function	**Results Cut-off Point**	**Term Expansion Level**	**Ignore Duplicates**
idf	10	Question Only	Yes
Resolve Pronouns	**Translate Terms**	**Similarity Measure:**	**Re-rank**
Yes	Yes	LESK	No

The manual processing for the randomly selected question and answers involved the following procedures:

- Question Enhancement with QurTerms and WordNet synonyms
- Answers Enhancement through Pronouns Addition using QurAna
- Removal of Punctuation from both Question and Answers
- Chunking of both Question and Answers
- Stopwords Removal from both Question and Answers
- Stemming of both Question and Answers
- Terms Generation from both Question and Answers

Once terms of both question and answers were generated, weights of individual terms within the selected answer were calculated. The weight of the answer was calculated by aggregating the weights of terms matching with the terms in the question. The potential answer with highest score was selected as *Answer*. Detailed evidence of testing procedure is available in Appendix 6.

Finally, output of these manual steps was verified with the output produced by the PARMS evaluation tool in order to verify validity of results produced by the PARMS evaluation tool. The following table shows the final outcome of the experiment:

Item	Result Produced By		Remarks
	Manual Process	**PARMS Evaluation Tool**	
Question terms	money, retribut, qisa, innoc, payment, blood, jurisprud, ransom, requit, law, equit, kill, equal, compens, punish, retali	money, retribut, qisa, innoc, payment, blood, jurisprud, ransom, requit, law, equit, kill, equal, compens, punish, retali	Result matched. Output of manual process is re-ordered for ease of comparison

Item	Result Produced By		Remarks
	Manual Process	**PARMS Evaluation Tool**	
Answer (002:178) terms	bloodwit, slain, lord, guardian, exceed, chastis, prosecut, prescrib, remiss, allevi, payment, brother, good, manner, kill, free, transgressor, null, matter, slave, pain, retali, made, murder, merci, aggriev, limit, usag, femal	bloodwit, slain, lord, guardian, exceed, chastis, prosecut, prescrib, remiss, allevi, payment, brother, good, manner, kill, free, transgressor, null, matter, slave, pain, retali, made, murder, merci, aggriev, limit, usag, femal	Result matched. Output of manual process is re-ordered for ease of comparison
Answer (002:178) weight (idf)	27.88257768	27.8825776990908	Result matched. Weight of individual matching terms has also been verified.
Answer (005:095) terms	deed, ihram, offer, person, lord, equival, wear, hunt, state, retribut, mighti, feed, pardon, judg, unwholesom, kaaba, penalti, poor, result, kill, pilgrimag, compens, fast, inflict, tast, expiat, anim, return, intent, cattl, game, brought	deed, ihram, offer, person, lord, equival, wear, hunt, state, retribut, mighti, feed, pardon, judg, unwholesom, kaaba, penalti, poor, result, kill, pilgrimag, compens, fast, inflict, tast, expiat, anim, return, intent, cattl, game, brought	Result matched. Output of manual process is re-ordered for ease of comparison
Answer (005:095) weight (idf)	25.42314606	25.4231460804535	Result matched. Weight of individual matching terms has also been verified.
Top Ranking Answer	002:178	002:178	Result matched. Manual result is out of two answers

Summary

IR and NLP tools and techniques can be combined to implement a QA system. The PARMS methodology has been presented in this chapter, which utilises these concept and methods to implement a QA system. This is capable of mapping a user question, expressed in natural language, to most relevant verses in the Quran translations. The PARMS methodology is implemented as a PARMS QA System and an evaluation framework is developed to analyse the effect of different concepts and methods used within the PARMS methodology.

Chapter 4 – PARMS Evaluation System

In order to evaluate a QA System for the Quranic QA, there is very little margin for error. This is because a wrong answer could provide misleading information that is not acceptable for religious reasons. This chapter introduces the PARMS Evaluation System which comprises the PARMS Evaluation Framework and the PARMS Model Question Set to test the PARMS QA System.

The PARMS (Parallel-corpus Multi Stream) QA System has been developed and implemented as a proof of concept for this research. It has been developed with the intention to evaluate the proposed PARMS Methodology. PARMS accepts English-language questions as an input. The output of PARMS is not an 'answer' in the classic sense. PARMS provides answers in the form of reference numbers to Quranic verse/s. The PARMS-generated "answer" is a verse reference number. The first three digits of this number refer to a chapter number followed by the verse number from within that chapter. For example 003:161 refer to verse 161 in chapter 003. This reference number can then be mapped to the Arabic Quran or a translation of the Quran in any language - English for this research work.

The PARMS Evaluation System uses a batch of 100 model questions to test the PARMS methodology against the corpus using different parameters discussed later in this chapter. A set of documents (0…n) are retrieved against each question. An empty set indicates that the system was not able to answer this question at all. A set of (1…n) document is called a '*result'*. Each document within the *result* is called an '*Answer*'. The *answers* within *result* are ranked according to the degree of relatedness to the question as determined by the system. The underlying rationale is that the higher the rank the higher the degree of relatedness and vice versa. Each *answer* within the *result* is correct if its reference matches one of the '*correct answers,* a pre-defined set of document references identified by the experts'. The query process is considered a '*success'* if 1 or more *correct answer* references, exist in the *result* and the *result* is declared as correct. If none of the *correct answer* references exist in the *result*, the query process is declared a *'failure'* and the *result* is deemed incorrect. Different statistical measures regarding the total number of results retrieved, total number of answers retrieved, and rank of each retrieved answer are also calculated.

This chapter content has been ordered such that the first section introduces the PARMS Model Question Set. The next sections provide details of the PARMS Evaluation Framework and details of the evaluation parameters used in the PARMS Methodology with their settings for groups of experiments. The last section describes the evaluation metrics and the tools used to detail the results of the experiments presented in the next chapter.

4.1 PARMS Model Question Set

The PARMS approach has been evaluated using a corpus of 100 Question Answer pairs collected from different original resources. These model QA pairs use natural language, English in this case. These sample questions exemplify 'real questions' as they had been asked by 'real people', and answers are based on discussions (Fatwas) provided by real domain experts – 'Muftis'. These have been designed for multiple purposes and have been gathered from multiple sources such as:

- Extracted from Authentic online scholarly sources
- Acquired from Educational Quranic QA websites
- Collected from Potential End users, i.e. Domain Users
- Derived from all the above – Questions designed to meet specific experiments

The following sections provide further details of the rationale, sources used, methodology adopted for gathering these sample questions and finally the design and representation of the Model Question Set as used by PARMS QA System. The Model Question Set is available in appendix 4.

4.1.1 Rationale

Our rationale in collecting questions and designing the PARMS Model Question set was the fact that the PARMS QA Methodology has been developed to deal with natural language (English) questions; hence the evaluation of the methodology should also be conducted using samples of natural language questions. Usually QA systems can be evaluated against a sample set of questions provided by organisations providing and supporting the infrastructure required for research in the particular area. For example TREC (NIST 1992) provides appropriate evaluation techniques and sample question sets for Question Answering on newswires, blogs and similar data. There is also work available on recommendations on how to design a reusable QA test collection (Voorhees and Tice 2000, Lin and Katz 2006). Unfortunately there is no benchmark set of test questions available for evaluating research on English translations of the Quran. There is evidence of some question sets produced in the past but either they are not sufficient, not appropriate or not available for evaluating the PARMS QA Methodology. Hence there was a need to develop a suitable Model Question Set.

To develop that Model Question Set, one approach would have been to accumulate this set of questions using offline methods, e.g. interviewing, question answering, conversations with real-world users and domain experts. A pilot run of this approach was conducted with

the support of the Islamic Society of Huddersfield University (ISOC 1980) and other resources, e.g. subject experts in the local community. The pilot produced a few useful questions which have been recorded for reference. These sample questions have been tagged as "Offline User Queries" in the Model Question Set and further details are provided in section 4.1.3.

The offline approach was laborious, and time consuming, and required a huge amount of additional work. A few social and religious problems with this approach were also noticed. Hence it was not adopted for any further work. It was decided to use existing QA resources on Islamic questions to develop an evaluation question set. There are published resources available offline, e.g. dedicated sections in Islamic newspapers and online, e.g. Islamic websites and forums. The online resources were selected as they were readily available and more accessible. Also these websites have a huge volume of QA pairs, answered by domain experts in response to real user questions. The questions are in natural language (English) and have been gathered over a long period of time, e.g. Islam QA service has questions collected since 1997. Further details of these collection resources and methods adopted are provided later in this chapter.

4.1.2 Collection Resources

There are several Question Answering resources available online where scholars (Imams, Muftis) and other domain experts provide answers to user questions asked by real world users; for example Islam Question Answering (Al-Munajjid 2013), Study the Quran (Ali 2013), Islam City (Imam 2013), Quran Al-Islam (Salhey 2013), Islam Web (Mufti 2013), Qibla Answers (Keller 2013), On Islam (Scholars 2013), Understanding Islam (Amjad 2001), Alim (Shah 2010), Studying Islam (Rahim 2007), and Renaissance Journal (Sheikh 2006).

A collection of Model QA pairs is influenced by several aspects of these available resources. Most importantly as religious information dissemination is involved, the source ought to be authentic, freely available and presented in English. Also a resource was preferred if the source has previously been used in related research. These filters were applied to assess the suitability of the resource for the PARMS QA System evaluation. The questions included in the Model Questions Set have mainly been taken from the following websites:

- Islam Question & Answer, i.e. www.IslamQA.info
- Study The Quran, i.e. http://studythequran.com
- IslamiCity, i.e. www.islamicity.com
- Islam Web, i.e. www.islamweb.net
- On Islam, i.e. http://onislam.net

- Qibla Answers, i.e. http://spa.qibla.com
- Understanding Islam, i.e. http://forums.understanding-islam.com/

Studying Islam (Rahim 2007), Alim (Shah 2010) and Quran Al-Islam (Salhey 2013) websites have also proved very useful in the design process of the Model Questions Set. QA services available through these websites address topics including Islamic theology and history, Islamic finance and social law, Islamic code of conduct (Fiqh) and jurisprudence, Islamic faith and belief, Islamic rituals and worship, social manners and etiquettes. A brief introduction to the websites used for question selection has been provided in section 2.3. Details of question selection procedure have been provided further.

4.1.3 Question Answer Pair Classification

The Model Question Answer pairs belong to different categories as discussed below:

Questions

(i) Frequently Asked Questions

Frequently Asked Questions (FAQ's), as the name suggests, are those questions that have been asked about a particular topic repeatedly. In practice, FAQ could refer to several types of question sets or information-disseminating articles (Hersch 1998). It is a common practice to define a section as a FAQ if it provides a list of questions that have either been frequently asked or are anticipated as possible questions; for example FAQ lists on Google, Yahoo or Islamic FAQ websites (Soltan 2012) (Marican 2013) (Shah 2010).

```
<Question id="45" enabled="true">
  <Text> In how many days did Allah create the earth?</Text>
  <Answers>
    <Answer rank="1">041:009</Answer>
  </Answers>
  <Topic>Faith:Belief</Topic>
  <SourceContext>
    <Source>http://studythequran.com/ShowQandA.php?qa_id=326</Source>
    <DateTaken>2013-03-22</DateTaken>
    <QuestionDate> Unknown </QuestionDate>
    <QuestionType> FAQ </QuestionType>
  </SourceContext>
</Question>
```

Figure 4.1: FAQ Type PARMS Model Question QID#45

We have used the term 'FAQ' to reference a set of questions anticipated about different topics discussed in the Quran as in the Q&A section on the "Study the Quran" website (Ali

2013). These questions are structured such that they use a natural language question text which is answered with verses from the English translation of the Quran. The PARMS QA system borrows these questions from their source in their original form without any pre-processing. An example of a FAQ question is given in figure 4.1 showing QID#45.

(ii) Online User Queries

In the context of this thesis, the term *Online User Queries* refers to questions that are available online and have been asked by real-world users of actual Islamic scholars. These questions have been taken from online resources listed in the previous section and are used in their original form without any significant pre-processing of the question text. But proof reading of these questions has been performed in order to remove any typos and simple errors. The answers to such question are at times also pre-processed. Details of this processing are provided later in the Answers section. The following figure provides an example model question QID#18 as taken from the IslamQA website:

```
<Question id="18" enabled="true">
  <Text>After I divorce my wife, is her mother regarded as a mahram?</Text>
  <Answers>
    <Answer rank="1">004:023</Answer>
  </Answers>
  <Topic>Islamic Ruling:Unmarriageable Kin</Topic>
  <SourceContext>
    <Source>http://islamqa.info/en/ref/95757</Source>
    <DateTaken>2012-11-22</DateTaken>
    <QuestionDate> Unknown </QuestionDate>
    <QuestionType> Online User Query </QuestionType>
  </SourceContext>
</Question>
```

Figure 4.2: Online User Query Type PARMS Model Question QID#18

(iii) Offline User Queries

An *Offline User Query* can be described as a query that has been collected during the pilot run of offline question collection. Such questions have been collected during question answering sessions with real-world users and the answers to these questions have been determined by real-world scholars during question answering sessions arranged with them. The following figure is an example of a PARMS Model Question QID#69 that has been collected using this process:

```
<Question id="69" enabled="true">
  <Text>What sort of volume I should recite Quran in when offering my Namaz?</Text>
  <Answers>
    <Answer rank="1">017:110</Answer>
  </Answers>
  <Topic>Rituals:Prayer</Topic>
  <SourceContext>
    <Source><![CDATA[ Offline ]]></Source>
    <DateTaken>2012-07-14</DateTaken>
    <QuestionDate> Unknown </QuestionDate>
    <QuestionType> Offline User Query </QuestionType>
  </SourceContext>
</Question>
```

Figure 4.3: Offline User Query Type PARMS Model Question QID#69

(iv) Derived User Queries

As the name suggests this type of user queries are in fact in the pre-processed form of online user queries. The available online questions are mostly in too much detail as the inquirers tend to provide background knowledge to set the context in which their actual question lies. This makes the question text unclear and noisy. Another difficulty with these questions is that they are usually double-barrelled. There is more than one question asked at one event. These questions need pre-processing to determine the actual text that contains the question. The following figure shows a model question QID#24 that has been derived from a double-barrelled Online User Query:

```
<Question id="24" enabled="true">
  <Text>What is the alternate if water is not found for ablution?</Text>
  <Answers>
    <Answer rank="1">005:006</Answer>
  </Answers>
  <Topic>Islamic Ruling:Ablution</Topic>
  <SourceContext>
    <Source>http://www.studythequran.com/ShowQandA.php?qa_id=1</Source>
    <DateTaken>2012-11-22</DateTaken>
    <QuestionDate> Unknown </QuestionDate>
    <QuestionType> Derived User Query </QuestionType>
  </SourceContext>
</Question>
```

Figure 4.4: Derived User Query Type PARMS Model Question QID#24

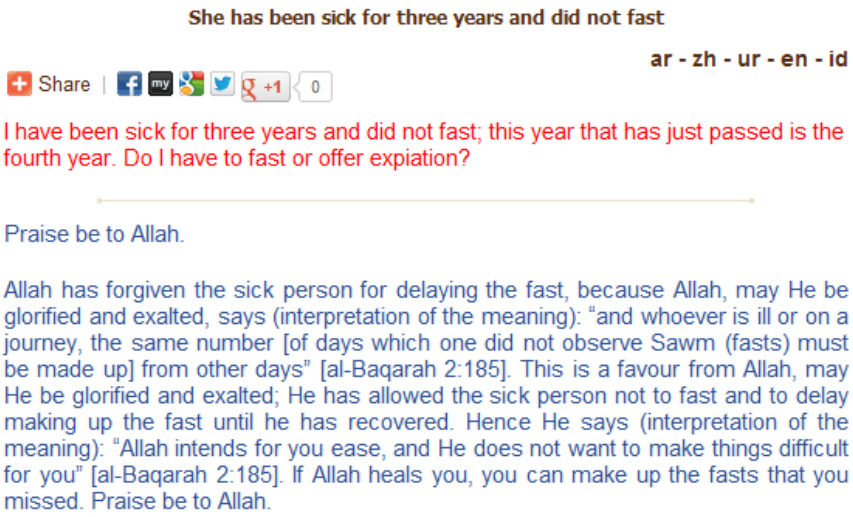
Answers

It is important to mention that, in most cases, the answers to the questions included in the PARMS Model Question Set are not the original answers but are extracts from these. Based on this fact the answers can be classified in two broad categories as follows:

(i) Expert Answers

Expert Answers are those answers that have been provided by the domain experts, i.e. the Islamic Scholars. Mostly, scholars use references directly taken from the Quran in the form of whole or partial Quranic verses. Consider the QA pair taken from IslamQA website (fatwa# 128809) presented in the quotation (figure 4.5). In this example, the scholar refers to verse 002:185 in logical parts in making his argument while answering the actual user question.

In this sample QA pair 002:185 is directly related to the question. Hence it was imported into the Model Question set in its original form without any pre-processing. But most user questions are not as straight-forward as the above example. Also in most cases scholars use multiple references to Quranic verses to provide an answer. They usually refer to other


She has been sick for three years and did not fast

ar - zh - ur - en - id

Share | +1 0

I have been sick for three years and did not fast; this year that has just passed is the fourth year. Do I have to fast or offer expiation?

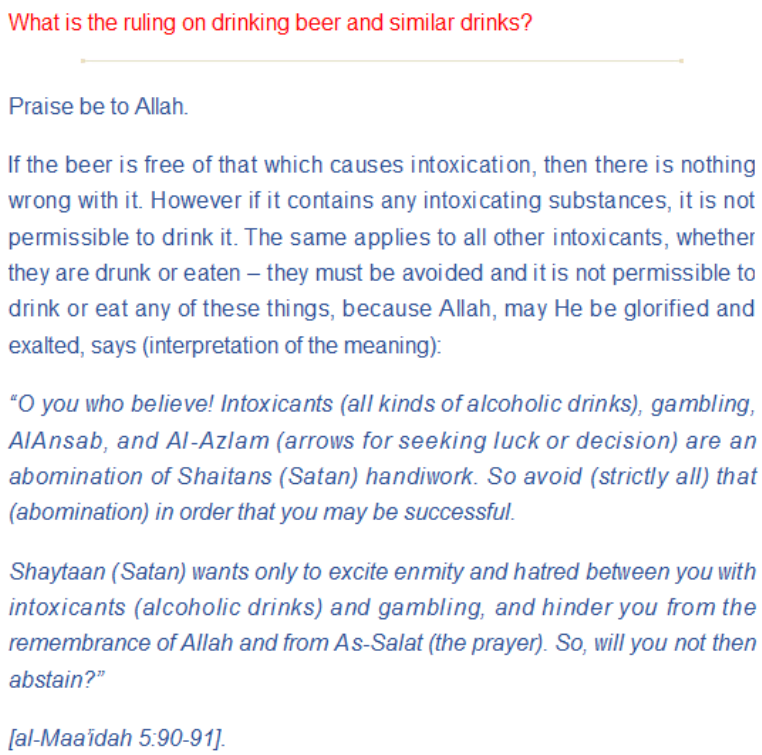
Praise be to Allah.

Allah has forgiven the sick person for delaying the fast, because Allah, may He be glorified and exalted, says (interpretation of the meaning): "and whoever is ill or on a journey, the same number [of days which one did not observe Sawm (fasts) must be made up] from other days" [al-Baqarah 2:185]. This is a favour from Allah, may He be glorified and exalted; He has allowed the sick person not to fast and to delay making up the fast until he has recovered. Hence He says (interpretation of the meaning): "Allah intends for you ease, and He does not want to make things difficult for you" [al-Baqarah 2:185]. If Allah heals you, you can make up the fasts that you missed. Praise be to Allah.

Figure 4.5: Question Answer sample from IslamQA.com

Islamic resources such as Hadith, Sunnah and Tafsir (exegesis) of the Quran. This happens because in answering a question scholar usually refers to more than one verse from the Quran alongside references from other resources.

Occasionally this leads to a situation when scholars support side arguments with the Quranic verse that are not related to the user question directly. Hence they add some 'noise' in the form of references to irrelevant verses from the Quran. Consider the example QA pair taken from IslamQA website (fatwa # 127020) presented in figure 4.6 below:



What is the ruling on drinking beer and similar drinks?

Praise be to Allah.

If the beer is free of that which causes intoxication, then there is nothing wrong with it. However if it contains any intoxicating substances, it is not permissible to drink it. The same applies to all other intoxicants, whether they are drunk or eaten – they must be avoided and it is not permissible to drink or eat any of these things, because Allah, may He be glorified and exalted, says (interpretation of the meaning):

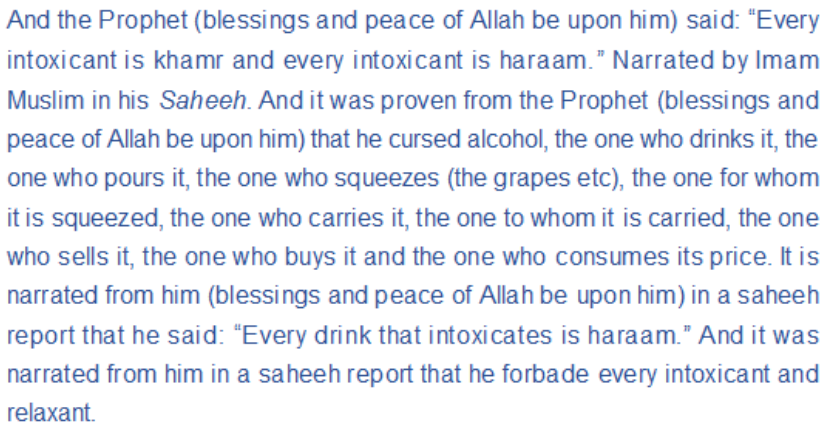
"O you who believe! Intoxicants (all kinds of alcoholic drinks), gambling, AlAnsab, and Al-Azlam (arrows for seeking luck or decision) are an abomination of Shaitans (Satan) handiwork. So avoid (strictly all) that (abomination) in order that you may be successful.

Shaytaan (Satan) wants only to excite enmity and hatred between you with intoxicants (alcoholic drinks) and gambling, and hinder you from the remembrance of Allah and from As-Salat (the prayer). So, will you not then abstain?"

[al-Maaïdah 5:90-91].

Figure 4.6: Question Answer sample from IslamQA.com – (a)

Further to support the argument a scholar may make reference to other Islamic resources such as Hadith and Tafsir in answering this question. This can be observed in figure 4.7:



And the Prophet (blessings and peace of Allah be upon him) said: "Every intoxicant is khamr and every intoxicant is haraam." Narrated by Imam Muslim in his *Saheeh*. And it was proven from the Prophet (blessings and peace of Allah be upon him) that he cursed alcohol, the one who drinks it, the one who pours it, the one who squeezes (the grapes etc), the one for whom it is squeezed, the one who carries it, the one to whom it is carried, the one who sells it, the one who buys it and the one who consumes its price. It is narrated from him (blessings and peace of Allah be upon him) in a saheeh report that he said: "Every drink that intoxicates is haraam." And it was narrated from him in a saheeh report that he forbade every intoxicant and relaxant.

Figure 4.7: Question Answer sample from IslamQA.com – (b)

Lastly a scholar may refer to some more Quranic verses in a commentary that support the concept of *'Piety'* and attributes of *'Believers'*. It can be observed from the quotation image (c) below that these Quranic verses do not relate to the user question directly but have been used in connection to later arguments.

What the Muslim should do is avoid all intoxicants and beware of them. The one who does any of that must give it up and hasten to repent to Allah, may He be exalted, from that, as Allah, may He be exalted, says (interpretation of the meaning):

"And all of you beg Allah to forgive you all, O believers, that you may be successful"

[an-Noor 24:31]

"O you who believe! Turn to Allah with sincere repentance!"

[at-Tahreem 66:8].

End quote.

Figure 4.8: Question Answer sample from IslamQA.com – (c)

A Question-Answer pair as shown above is not useful in its original form so some pre-processing was applied before use. In such cases we used personal judgment to extract the most relevant verse for given user question and nominated that verse as the best answer.

For instance in the above QA pair, the Quranic verses 005:090 and 005:091 are relevant in the context of the user question as compared to later referenced Quranic verses 024:31 and 066:008. Hence they were nominated as the potential answers to this user question.

(ii) Derived Answers

The PARMS Model Question Set also includes eight offline questions collected from users. The answers to these questions were derived using personal judgment and cross-verified by a scholar. Consider Model Question QID#69 as an example:

```
<Question id="69" enabled="true">
  <Text>What sort of volume I should recite Quran in when offering my Namaz?</Text>
  <Answers>
    <Answer rank="1">017:110</Answer>
  </Answers>
  <Topic>Rituals:Prayer</Topic>
```

Figure 4.9: PARMS Model Question QID#69

The answer to this question has been inferred. As the first step the concept behind the question is identified, which in the case of QID#69 is *'Prayer/Salaat/Salah'*. Furthermore this question relates to the *'volume of voice'* during prayer. This concept is looked up in the *Quran Subjects* feature of Quran Al-Islam website and *Al-Quran Subject Index* of Alim.org

website. The concept *'Prayer/Salaat/Salah' has* been discussed in the Quran Subject under the topic *Acts of Worship*. This can be observed from the website feature image below:

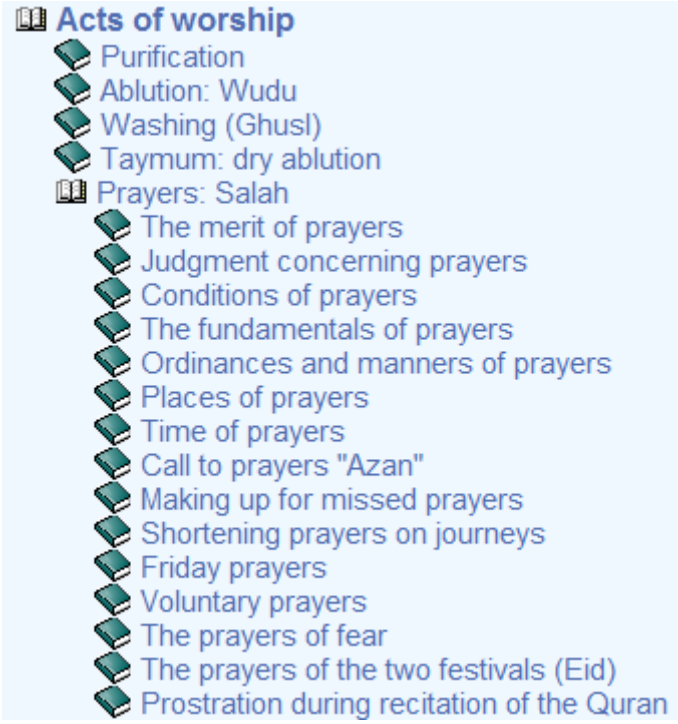


Figure 4.10: Quran Subjects feature of Quran Al-Islam website

But there is a reference to the *'volume of voice'* during prayer in the list of sub-topics discussed in the *Quran Subjects* feature of the Quran Al-Islam website. The *'Prayer/Salaat/Salah'* concept is also available in the *Al-Quran Subject Index* of the Alim.org website and it indicates the Quranic verse 017:110 directly related to the *'volume of voice'* during prayer. Hence the Quranic verse 017:110 is derived as the answer verse to question QID#69. This can be observed from the website feature image below:

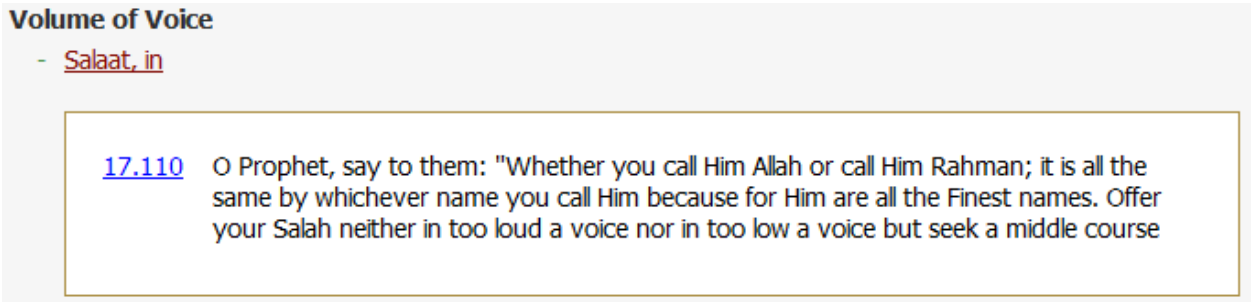


Figure 4.11: Al-Quran Subject Index of Alim.org website

Generally it was possible to find answers to offline questions using the same method but there are other minor investigation strategies that were used to find the finer details of

concepts. Also reverse engineering was adopted to identify question answer pairs in certain situations. For instance a topic of interest was searched in the *Question Categories* list of the Studying Islam website and adopted to be included in the PARMS Model Question Set.

4.1.4 Statistical Information:

Questions in the PARMS Model Question set have been grouped based on the Type and the Topic of the question.

(i) Type of Question

The PARMS Model Question Set contains questions of different types, i.e. FAQ, Online User Query, Offline User Query and Derived User Query. Details of these categories have been provided in section 6. The following table shows the statistics for these groups:

Question Type	No. of Questions
Online User Query	48
Offline User Query	8
Derived User Query	4
FAQ	40
Total:	100

Table 4.1: Statistics for Question Types in PARMS Model Question Set

(ii) Topic of Question

Question answer pairs in the PARMS Model Question Set have been grouped based on the common topic and concepts under question. These groups are listed in the table below. There is a variety in choice of questions to ensure several topics of interest are included in the PARMS Model Question Set. There are multiple questions in some topic groups whereas other groups contain as few as only two questions. Table 4.2 shows the statistics for these groups.

Code of Conduct for example encompasses questions on privacy, slavery, adoption, breastfeeding, charity, divorce, dowry, extravagance, family, inheritance, mahr, mosque, orphans, waiting period for widows and will. The question group Faith contains fourteen question and all of these are about belief in Islam. The Question Set includes four questions about the history of prophets. The Question set on Islamic Rulings include questions about ablution, abortion, adultery, business, making contract, corruption, debt, equality, food, haram, qisas, unmarriageable kin, usury, veiling and the concept of zihar. Also two model

Question Topic	No. of Questions
Code of Conduct	25
Faith	14
History	4
Islamic Ruling	31
Rituals	17
Social Norms	7
Quran	2
Total:	100

Table 4.2: Statistics for Question Topics in PARMS Model Question Set

questions ask about the Quran as a book. Rituals embrace questions on rules about fasting, pilgrimage and prayer. Lastly the Social Norms group includes questions on topics like effect of backbiting, forgiveness, mannerism, taking and keeping oath, trust and dealing with non-Muslims.

4.1.5 Representation

Question-Answer pairs in the PARMS Model Question Set are represented using XML format. This ensures the benefits of the XML technology as well as several other advantages like readability, reusability, etc. The XML schema used to represent the PARMS Model Question Set is available in Appendix 3 for further reference. The sample question below shows the specific information stored for each model question using the XML template:

```
<Question id="99" enabled="true">
  <Text> In what orders seven heavens have been created by Allah?</Text>
  <Answers>
    <Answer rank="1">067:003</Answer>
    <Answer rank="2">071:015</Answer>
  </Answers>
  <Topic>Faith:Belief</Topic>
  <SourceContext>
    <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=364 ]]></Source>
    <DateTaken>2012-11-22</DateTaken>
    <QuestionDate> Unknown </QuestionDate>
    <QuestionType> Derived User Query </QuestionType>
  </SourceContext>
</Question>
```

Figure 4.12 PARMS Model Question Sample

4.2 Evaluation Parameters

PARMS Evaluation Framework provides a configurable set of parameters which makes it possible to test the effect of different components and processes on the system results. The software developed for the investigation facilitates changing these parameters through a user-friendly interface. The Screenshot presented in figure 4.13 details the list of parameters that are supported in the experimental environment:

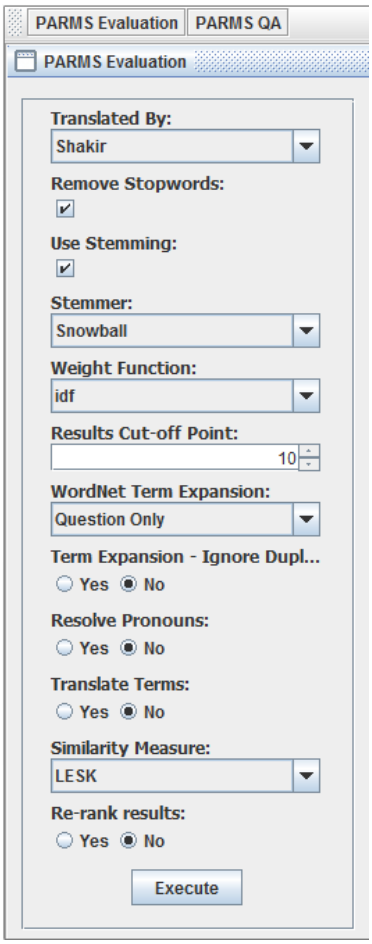


Figure 4.13 PARMS Evaluation System Parameters

Multiple sets of experiments were carried out to investigate the success variation of the PARMS QA Methodology when tested under different parameters. A brief description for each of these parameters is discussed below:

(i) Translated By

PARMS QA System allows access to ten different translations of the Quran as discussed in chapter 2 section 3. The PARMS Evaluation system makes it possible to use these translations in three modes as follows:

A – Single Author Translation Mode

This mode allows using one translation at a time for a particular experiment. This facility is available for all ten translations of the Quran embedded within the PARMS QA System. As the Quran has 6236 verses, if an experiment is run in this mode, it means that 6236 Quranic verses are under consideration as potential answer.

B – All (Parallel) Mode

This mode refers to using '*All (Parallel)'* ten translations in parallel to each other where each single translation stands as an independent entity. An experiment run in this mode means all ten translations for each verse are taken into consideration as a potential answer separately. For example the following is the *All (Parallel)* mode of translations for the Quranic verse 001:003.

001:003 Khan	*The Most Beneficent, the Most Merciful.*
001:003 Maulana	*The Beneficent, the Merciful,*
001:003 Pickthal	*The Beneficent, the Merciful.*
001:003 Rashad	*Most Gracious, Most Merciful.*
001:003 Sarwar	*The Beneficent, the Merciful*
001:003 Shakir	*The Beneficent, the Merciful.*
001:003 Sherali	*The Gracious, the Merciful.*
001:003 Yusufali	*Most Gracious, Most Merciful;*
001:003 FaridulHaq	*The most Affectionate, The Merciful.*
001:003 HilaliMohsin	*The Most Beneficent, the Most Merciful.*

As evident from above example, each translation has an equal potential to be selected as an answer to the question. These verses do not contribute to each other's weight. Rather this mode makes similar verses contenders to each other for selection as an answer. The number of Quranic verses in this mode can be calculated as:

6236 x 10 = 62360 Quranic verses

This means that each verse from each of the ten Quranic translations keeps its identity as an individual verse. Each of these 62,360 Quranic verses is then considered as a potential answer.

C – Parallel-Corpus Mode

All (Merged) is a parallel-corpus mode of translations that refers to a combination of the Quranic verses across ten translations in a unique way. In this mode all corresponding verses from all different translations are merged and diffused to form an entire new verse representation. For example the *Parallel-Corpus* mode of translation for the above Quranic verse 001:003 will yield the following merged verse structure:

> *The Most Beneficent, the Most Merciful, The Beneficent, the Merciful. The Beneficent, the Merciful. Most Gracious, Most Merciful. The Beneficent, the Merciful. The Beneficent, the Merciful. The Gracious, the Merciful. Most Gracious, Most Merciful. The most Affectionate, The Merciful. The Most Beneficent, the Most Merciful.*

It is important to note that this merged document is no longer a Quranic verse. This document is given the same verse number as the original verse for reference during processing. However if this is selected as an answer then this reference number is cross-referenced to any particular translation to retrieve the verse (i.e. answer) in its actual form.

Unlike *All (Parallel)* mode, in *Parallel-Corpus* mode each of ten *Single Author Translations* contributes its weight to the final *documen*t during the merger process. Hence the *Parallel-Corpus* mode of translations produces 6,236 *documents* for 6,236 Quranic verses where each *document* carries the summed weights of all ten translations for that particular Quranic verse.

(ii) Remove Stop Words

The PARMS Evaluation system provides an option to either use or not use a stopword list for a particular experiment. Currently the PARMS system uses only the stopword list provided by Stanford University. A few preliminary experiments were carried out using other stopword lists too but there was no significant difference in using another list.

It is believed that PARMS can benefit from having a dedicated stopword list designed to suit the Quranic dataset. This is a tedious task in itself and requires lots of research in the area, so is beyond the scope of this project. A quick-fix solution could be to study the PARMS system output using stopwords lists trained on other religious data such as the Bible and adopt these lists to suit Islamic stopword terms.

(iii) Use Stemming

This option helps decide whether or not to use stemming. If selected, it will stem different terms while creating the control vocabulary for both documents and the question.

(iv) Stemmers

If stemming is to be used for a particular experiment, the PARMS Evaluation system allows a choice between three stemmers, namely Snowball Stemmer, Porter Stemmer and Pling Stemmer. Algorithmic details of these stemmers are provided in section 3.1.4.

(v) Weight Functions

The PARMS Evaluation system facilitates term weighting using four weight functions namely Frequency of terms, Inverse document frequency count, Vector state model and OKAPI model. Choice between these for a particular experiment has implications for the results. These weighting functions are discussed in detail in section 3.1.2.

(vi) Results Cut-off Point

Results Cut-off point is the measure of window size for top retrieved answers. PARMS for instance, could retrieve over a few thousand relevant Quranic verses in initial processing, which are narrowed done after going through other filters to just over a few hundred. But the retrieved verse is considered correct only if it falls within the number indicated as the Result Cut-off point.

If the Results Cut-off point is set to window size 10, then the top ten retrieved verses will be considered as correct while if this window size is brought down to 1, the retrieved verse will have to appear as the top answer to qualify as the correct answer. The PARMS Evaluation system can be configured to certain experimental needs using this option.

(vii) WordNet Term Expansion

The PARMS Evaluation System can be configured to use the WordNet (Miller, Beckwith et al. 1990, Miller 1995, Fellbaum 1998, Pedersen, Patwardhan et al. 2004) ontology for term expansion. This term expansion can be applied at four different levels within the PARMS Methodology. These Options are as follows:

A – Questions Only

This option, if chosen, expands only the query terms using the WordNet ontology.

B – Documents Only

If this option is chosen, it allows expansion of only the document (Quranic verses) terms using the WordNet ontology.

C – Both (Questions and Documents)

This option expands both Query terms and document terms using the WordNet ontology.

D – None

If an experiment does not require any term expansion either for query or document, then this option is the appropriate choice for the experiment.

(viii) Ignore Duplicate Terms

Term expansion using the WordNet ontology at times causes duplication of terms. This can lead to additional weight for a particular term or even to noise in certain cases. Given this situation it is necessary to keep this duplication under control depending on the needs of the experiment. PARMS provides an option to choose either to keep the duplicated terms or ignore them for certain experiments.

(ix) Resolve Pronouns

The PARMS Evaluation System facilitates the use of the QurAna ontology through a user-friendly interface. The QurAna ontology is used to resolve pronouns in Quranic verses only if this choice has been enabled. Details of the QurAna ontology and how pronoun resolution procedures are adopted in the PARMS methodology are presented in earlier chapters in relevant sections.

(x) Translate Terms

The QurTerms ontology is activated to translate Islamic and Quranic terms if a choice has been made using this option. Details of the QurTerms ontology and translation algorithm embedded within the PARMS methodology are presented in earlier chapters in relevant sections.

(xi) Similarity Measure

The PARMS Evaluation System provides a facility to run experiments under two different similarity measures. The Cosine Similarity Measure is the first measure that allows a calculation of similarity between the query and document terms using the traditional Cosine method. The LESK Algorithm has been adopted to calculate similarity based on word overlap between the query and document terms. Theoretical details of these measures have been given in earlier chapter 3. Limitations and usefulness of both of these are discussed later in this chapter.

(xii) Re-Rank Results

Re-ranking is performed using the QurSim ontology. The PARMS Evaluation System allows for configuring of the experimental environment to either perform re-ranking or not. The theoretical details of the QurSim ontology are provided in chapter 2. Details of the PARMS method of re-raking answers using the QurSim ontology are discussed in Chapter 3. The evaluation results of the PARMS methodology using re-ranking are reported later in this chapter.

Some preliminary experiments suggest that certain parameter settings provide improved results as compared to others. For example the Snowball Stemmer produced better results and the LESK Algorithm out-performed the Cosine Similarity method. Similarly it was also observed that re-ranking answers using the QurSim ontology deteriorates overall results. Detailed experiments conducted using these parameter settings are discussed in the following section, while the parameter settings that were unable to produce better results are also presented in a special cases section later in this chapter for the sake of completeness.

4.3 Parameter Settings

The PARMS Methodology has been tested under different experimental environment settings. For ease of readability and understanding certain parameters have been collated below as groups of parameters. This makes it possible to refer to certain group of settings while focusing on only some specific aspect of the whole PARMS QA System. These groups of parameters have been assigned a name for future reference as follows:

Simple Parameters Setting (SPS)

This group of parameters is termed as 'simple' due to the simple settings used for these experiments. Simple Parameter Settings (SPS) include the following parameter settings for the experimental environment:

Translation	All 12 options	WordNet Expansion	None
Remove Stopwords	No	Ignore Duplicate Terms	None
Use Stemming	No	Resolve Pronouns	No
Stemmer	None	Translate Terms	No
Weight Function	$f_{d(t)}$	Similarity Measure	LESK
Results Cut-off Point	1, 5, 10	Re-rank Results	No

Table 4.3: Simple Parameter Settings

Basic Parameters Setting (BPS)

The Basic group of parameter settings is designed primarily to test Basic Information Retrieval parameters. Basic Parameter Settings (BPS) include the following parameter settings for the experimental environment:

Translation	All 12 options	WordNet Expansion	None
Remove Stopwords	Yes	Ignore Duplicate Terms	None
Use Stemming	Yes	Resolve Pronouns	No
Stemmer	Snowball	Translate Terms	No
Weight Function	idf	Similarity Measure	LESK
Results Cut-off Point	1, 5, 10	Re-rank Results	No

Table 4.4: Basic Parameter Settings

Enhanced Parameters Setting (EPS)

The Basic group of parameter settings is designed primarily to test enhanced NLP parameters. The Enhanced Parameter Settings (EPS) includes following parameter settings for the experimental environment:

Translation	All 12 options	WordNet Expansion	Questions Only
Remove Stopwords	Yes	Ignore Duplicate Terms	Yes
Use Stemming	Yes	Resolve Pronouns	Yes/No
Stemmer	Snowball	Translate Terms	Yes/No
Weight Function	idf	Similarity Measure	LESK
Results Cut-off Point	1 through to 10	Re-rank Results	No

Table 4.5: Enhanced Parameter Settings

4.4 Evaluation Metrics

The PARMS Evaluation System uses the most common evaluation metrics in Information Retrieval and Question Answering to evaluate the effectiveness and correctness of PARMS QA System results. An answer is accepted as correct if it is a *relevant verse* to the Question. The judgment of verse relevancy to the question is assumed to be within the borderline meaning of *'aboutness'* and *'appropriateness'* of the verse in the context of the question (Sultan 2011) (Rusbergen 1979).

The PARMS Evaluation system uses pre-determined relevant verse(s) for each question within the question set. These relevant verses have been identified by the experts answering the questions on their corresponding forums. Each question has one or more relevant verses and if any of these verses exists in the top *N* ranked answers retrieved by the system, then the answer is acknowledged as correct.

I – Accuracy

In the PARMS Evaluation system, the Accuracy measure is defined as the proportion of correctly retrieved results amongst the total number of questions being queried in the batch. This can be presented as follows:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{Number\ of\ correctly\ retrieved\ results}}{\boldsymbol{Total\ number\ of\ questions}}$$

To evaluate the PARMS Methodology, the PARMS Evaluation system uses a Model QA set comprising 100 questions. More information on this question set has been given earlier in this chapter.

II – Precision

The PARMS Evaluation System defines the Precision measure as the proportion of correct results amongst the total number of results retrieved by the system. This can be expressed as follows:

$$\boldsymbol{Precision} = \frac{\boldsymbol{Number\ of\ correctly\ retrieved\ results}}{\boldsymbol{Total\ number\ of\ retrieved\ results}}$$

or

$$\boldsymbol{Precision} = \frac{|\{\boldsymbol{relevant\ results}\} \cap \{\boldsymbol{retrieved\ results}\}|}{|\{\boldsymbol{retrieved\ results}\}|}$$

The aim of the PARMS methodology is to retrieve the correct answers as high as possible in the top ranked list of answers. The PARMS Evaluation System facilitates the measuring of precision at different levels on the ranked list. The PARMS Results cut-off point parameter permits configuring the measure of precision at different ranks. Mostly experiments are performed at three ranks, as in Sultan (2011).

(i) Precision at 1st rank – P@1

Precision measure remains the same as defined above for P@1. In fact standard precision is P@1 inherently. The P@1 measure takes only those potential answers that appear at 1st rank in calculating precision.

(ii) Precision at top 5 ranks – P@5

The P@5 considers the top 5 potential answers for each question when measuring the precision of the system. This implies that P@5 is Precision at n, where n = {1, 2, 3, 4, 5}.

(iii) Precision at top 10 ranks – P@10

Precision at top 10 ranks is a standard precision as in the case of P@1 and P@5. P@10 includes the top 10 potential answers for each question evaluation. P@10 is Precision at n where n = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}.

III – Recall

Recall, in the PARMS Evaluation System, is defined as the proportion of correctly retrieved answers amongst the total number of correct answers, pre-identified by experts, which exist within the corpus for a given question set. Recall can be represented as:

$$\boldsymbol{Recall} = \frac{\boldsymbol{Number\ of\ correctly\ retrieved\ answers}}{\boldsymbol{Total\ number\ correct\ answers\ within\ corpus}}$$

Or

$$\boldsymbol{Recall} = \frac{|\{\boldsymbol{relevant\ answers}\} \cap \{\boldsymbol{retrieved\ answers}\}|}{|\{\boldsymbol{relevant\ answers}\}|}$$

IV – Mean Average Precision

The Mean Average Precision (MAP) for the given question set is the mean of the average precision score for each question. Mathematically this can be presented as:

$$\boldsymbol{MAP} = \frac{\sum_{\boldsymbol{q=1}}^{\boldsymbol{Q}} \boldsymbol{AvgP(q)}}{\boldsymbol{Q}}$$

Where ***Q*** is the total number of questions in the question set, while ***AvgP(q)*** is the average precision of each question ***q***. ***AvgP(q)*** is the average of the precisons calculated at each point when a relevant answer is retrieved within each result for question **q**.

V – Mean Reciprocal Rank

The Mean Reciprocal Rank (MRR) for a given question set is the average of the reciprocal rank for each question. The reciprocal rank of a question response is the multiplicative inverse of the rank of the first correct answer.

$$\boldsymbol{MRR} = \frac{\mathbf{1}}{\boldsymbol{|Q|}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{|Q|}} \frac{\mathbf{1}}{\boldsymbol{rank_i}}$$

In the above equation Q is the total number of questions in the question set.

4.5 Results: Visual Representations

This section introduces the structures, means and tools used to represent the results of the experiments conducted. Different structures and means have been used to represent the results of different experiments. This is because certain experiments have been discussed in much more detail as compared to other less important experiments. The following are brief introductions. Further details relevant to specific experiments are provided within the discussion of those particular experiments.

I – Results Table

A sample of the tables devised to represent the experiments comprehensively is shown below. Each row of the table represents one whole experiment while each column represents the particular attribute being tested by a group of experiments. Each intersection cell describes a certain attribute of a particular experiment.

Translator	Total Questions	Total Correct Answers	Total Retrieved Results	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP	MRR
Maulana	100	156	98	34	37	0.34	0.3469	0.2372	0.2085	0.6148
Pickthal	100	156	98	36	41	0.36	0.3673	0.2628	0.1484	0.4147
Yusuf Ali	100	156	98	42	50	0.42	0.4286	0.3205	0.2194	0.5315
Sher Ali	100	156	98	45	49	0.45	0.4592	0.3141	0.2045	0.4498
Sarwar	100	156	98	43	48	0.43	0.4388	0.3077	0.2765	0.6582
Shakir	100	156	98	34	36	0.34	0.3469	0.2308	0.2169	0.6337
Rashad	100	156	98	37	39	0.37	0.3776	0.2500	0.2197	0.6012
Khan	100	156	99	45	51	0.45	0.4545	0.3269	0.2244	0.5058
Hilali & Mohsin	100	156	99	44	50	0.44	0.4444	0.3205	0.2263	0.5220
Farid-ul-Haq	100	156	99	43	48	0.43	0.4343	0.3077	0.2601	0.6292
All	100	156	99	51	59	0.51	0.5152	0.3782	0.3092	0.6359
All (Combined)	100	156	99	57	66	0.57	0.5758	0.4231	0.3343	0.5984

Table 4.6: Sample BPS Experiment Result for P@10

The first column in the table presents the translation mode used for a certain experiment. The second column is the total number of questions used, i.e. 100 model questions. Each

```
<Question id="09" enabled="false">
  <Text>What is the prescribed waiting period for a widow?</Text>
  <Answers>
    <Answer rank="1">002:234</Answer>
  </Answers>
  <Topic>Code of Conduct:Waiting Period</Topic>
  <SourceContext>
    <Source>http://www.studythequran.com/ShowQandA.php?qa_id=126</Source>
    <DateTaken>2012-11-22</DateTaken>
    <QuestionDate> Unknown </QuestionDate>
    <QuestionType> FAQ </QuestionType>
  </SourceContext>
</Question>
```

Figure 4.14: Sample Model Question - QID#09

question in the Model Question Set can have more than one correct answer. This can be better understood with an example question as shown in figure 4.14. The model question shown in the figure has only 1 possible answer as proposed by source scholars for the given question. But in model question ID# 50 presented in the next figure 4.15, scholars indicate 3 potential verses that provide supporting information for the answer.

```
<Question id="50" enabled="false">
  <Text>Is the doctors' knowledge of the fetus in the uterus complete?</Text>
  <Answers>
    <Answer rank="1">031:034</Answer>
    <Answer rank="2">013:008</Answer>
    <Answer rank="3">053:032</Answer>
  </Answers>
  <Topic>Faith:Belief</Topic>
  <SourceContext>
    <Source>http://islamqa.com/en/ref/1513</Source>
    <DateTaken>2012-11-22</DateTaken>
    <QuestionDate> Unknown </QuestionDate>
    <QuestionType> Online User Query </QuestionType>
  </SourceContext>
</Question>
```

Figure 4.15: Sample Model Question - QID#50

The PARMS Model Question Set comprising of 100 questions has 156 relevant correct answers as recommended by scholars. This is the number represented in the third column of the table. Both the 2nd and 3rd column are hidden in the table representations unless they are required for a particular experiment. This is because they do not serve much purpose as their value remains constant across experiments. The fourth column of the table gives information on the total number of answers retrieved regardless of whether they are correct or incorrect.

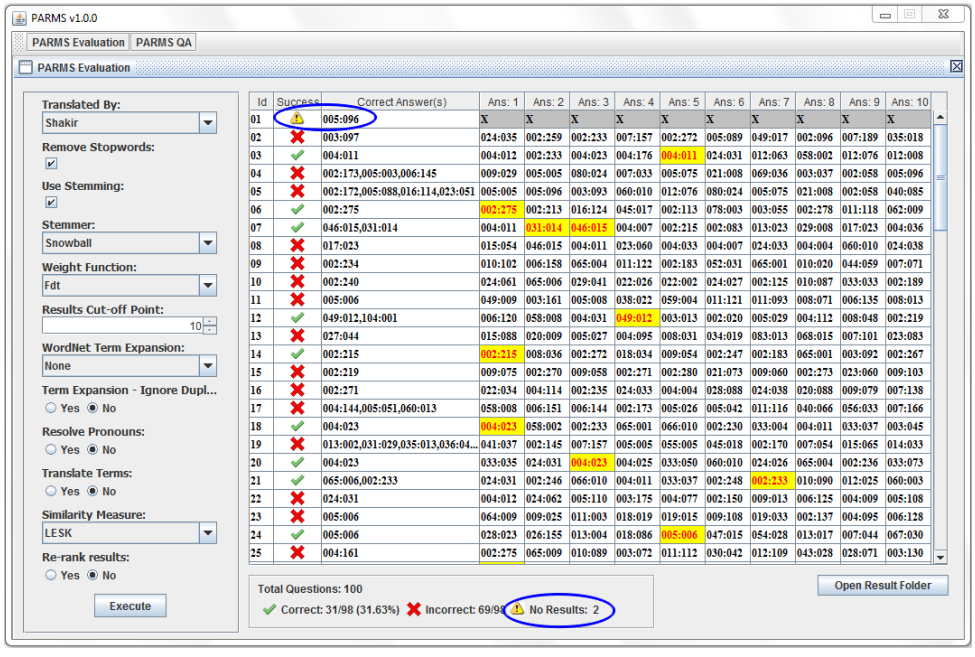


Figure 4.16: Screen Shot of Unsuccessful Results shown

There are certain cases where some model questions do not yield any answer using the PARMS Methodology. For example the above screen shot shows a case where the PARMS Methodology was unable to produce results for 2 model questions.

This means that results have been retrieved for only 98 questions out of 100 model questions. The fifth column provides information on the number of correct results produced by each experiment. This does not include results which have been retrieved but are incorrect. The next five columns provide accuracy, precision, recall, Mean Average Precision (MAP) and Mean Reciprocal Rank values for each experiment respectively. Details of the calculations used for these fields have been provided in section 4.4.

II – Bitmap Table

Bitmap Table is a term assigned to a tabular representation of experimental results. Results of the experiments require reporting on the following:

- Boolean values for a set of 100 question answer pairs – 1 stands for correct answer and 0 represents incorrect answer.
- Answers of above set of 100 question answer pairs when tested against 12 different data streams.
- Performance of each data stream as compared to other 9 data streams.

A bitmap table (sample below) has been used to depict results of individual translation mode for all 100 model questions at a glimpse. Each cell in the bitmap is an intersection of a translation mode and a specific question.

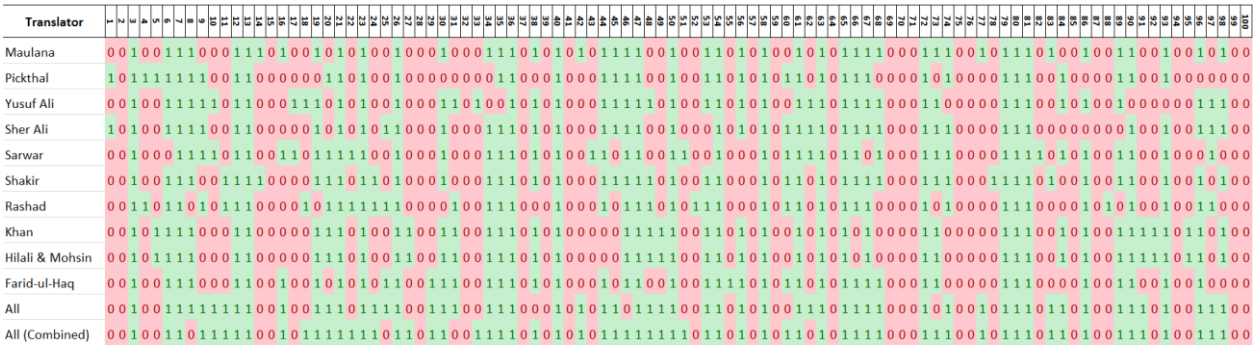


Figure 4.17: Sample EPS Bitmap Image for P@10

A value of *'1'* in table cell depicts *Success* of the answer retrieved whereas a value of *'0'* shows a *Failure*. These cells are also colour-coded for better readability.

III – Change Chart

A Change Chart is a visual aid devised to depict the effect of change during gradual processing of experiments. Different icons represent different aspects of change. For example the  icon represents positive change,  represents negetive change and  represents no change.

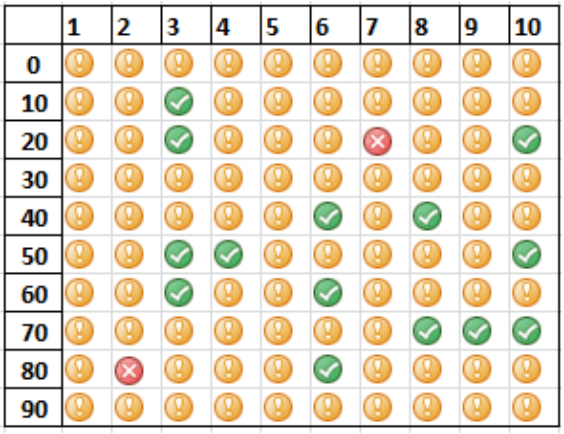


Figure 4.18: Change Chart Depicting Change of Results

Positive change occurs when a question has been successful as compared to its failure in the previous stage of the experiment. A negative change represents the fact that the successive stages are causing negative impact on the results, i.e. the correct answer is now incorrect.

This visual aid is used to present results of experiments that have gradual input and multiple stages, for instance when impact of a certain input needs to be studied as compared to the earlier input. This can be witnessed in detailed experiments presented in Chapter 5.

Chapter 5 – PARMS Evaluation Results

PARMS uses a combination of advanced IR methods and NLP techniques to extract *terms* from the query and individual documents (verses) respectively. Later these *terms* are augmented using general and domain specific ontologies. The resultant *terms* are then weighted using a weighting function. In the second stage, query *terms* are mapped to document *terms*, and the similarity between the query and each document is calculated by cumulative weights of the *terms* occurring both in query and document. Documents, now termed 'candidate answers', are then ranked according to their similarity scores. The candidate answers are further re-ranked using domain specific ontologies and top *N* re-ranked document(s) are declared the answer. The results of a batch of questions are collated and evaluated. Evaluation results are quantified using different standard statistical measures.

The aim of the evaluation is to test the proposed methodology in order to identify a combination of different parameters that produces the best results. The evaluation has been carried out to test the performance of domain-independent methods, i.e. without using domain-specific resources, and compare it to the results obtained using domain-specific knowledge with application to the Quranic dataset.

The first section of this chapter discusses the preliminary experiments which were conducted to observe the effects of certain PARMS Methodology parameters as compared to other parameters. These experiments also made it possible to determine which parameters provide significant results compared to others. The findings of these experiments are studied in much more detail to carry out further experimentation using only significant sets of parameter settings as discussed in section 4.3. These detailed experiments have been discussed in the third section of this chapter. Lastly, an overall analysis of results has been presented before the closing summary of this chapter.

5.1 Preliminary Experiments

Preliminary experiments have been carried out to study the effects of the evaluation parameters in general and identify success trends. The goal of these experiments is to determine which parameter values produce poor results and what parameter combinations produce successful results for the PAMRS method. The findings from this section have been used to inform later detailed experiments discussed in the next section of this chapter.

Different translations offered by the PARMS Methodology have been tested to identify the role of this parameter. Initially experiments were conducted on single-author translations. Later experiments used *All (Parallel)* and *All (Merged)* modes to identify the value of the parallel-corpus mode on individual translation mode. These experiments were run under simple parameter settings (SPS). The collated results of these experiments in order of the year of translation publication are provided in section 5.2.

5.1.1 Effect of Basic NLP Techniques

Initial experiments on basic NLP techniques including use of stopwords and stemming were conducted to observe the implication of these techniques on the PARMS Methodology. Use of stopwords did not show significant results. We believe that this insignificant contribution is down to the fact that so far no stopwords list has been developed which is suitable for English translations of the Quranic corpus.

The PARMS methodology uses an existing and popular English stopword list, produced by the Stanford University research group. This list fails to remove less informative words from this dataset. Related research work indicates that a few researchers have developed stopword lists for personal projects. But these are not available for public research. Also as they were developed for specific projects, they use specific languages, e.g. Malay and Arabic. We believe there is a need for a specialist stopword list for English translations of Quranic corpus. This can be seen as a potential research area to explore independently.

Stemming has been tested using different stemmers. These include the Snowball, Porter and Pling stemmers. Preliminary experiments were run on different variations of parameters to determine the best suited stemmer for the PARMS method. There was no significant difference found in results for each of these stemmers. The Snowball stemmer outperforms the other two when applied to the Quran translation authored by Shakir. This translation uses contemporary English. The Porter stemmer produces best results with the translation done by Pickthal, the oldest translation amongst those used by PARMS. The Pling stemmer generally performs the worse of the three but it produced the best results when the Inverse Document Frequency (idf) term weighting was used.

In later experiments, the addition of new parameters gradually established a general trend in results. Generally the Snowball stemmer preformed relatively better under basic parameter settings (BPS) as compared to other stemmers. Hence this stemmer was used to conduct further experiments using the enhanced parameter settings (EPS).

5.1.2 Effect of Term Weighting Functions

Preliminary experiments show that the Inverse Document Frequency (idf) function generally outperforms all other term-weighting functions. The idf function produces 63% correct results with Enhanced Parameter Settings (EPS) and 49% with Simple Parameter Settings (SPS). The Term Frequency (fdt) function produces the least correct answers, i.e. 15% with SPS. The fdt function is able to produce a maximum of 50% correct results with EPS whereas the vector state model (fdt x idf) achieves these 50% correct results using Basic Parameter Settings (BPS).

The fdt x idf function produces a maximum of 57% correct results when all (WordNet, QurAna, QurTerms) additional semantic knowledge is embedded in EPS. The idf function also produces 57% with limited (WordNet only) additional semantic knowledge using Basic Parameter Settings (BPS). The OKAPI model's best performing results reach 58% with EPS and are limited to 29% with SPS.

Analysing the above experiments, it is deduced that the Inverse Document Frequency (idf) function performs relatively better than other term-weighting functions. Hence idf has been used to conduct further detailed experiments discussed later in this chapter.

5.1.3 Effect of WordNet Ontology

The effect of the WordNet (Miller, Beckwith et al. 1990, Miller 1995, Fellbaum 1998, Pedersen, Patwardhan et al. 2004) ontology is visible as compared to the NLP techniques. But this effect is prominent when the WordNet ontology is used to process the user question, not the answer documents. Preliminary experiments show that user question terms expanded using the WordNet ontology help identify more correct answers as compared to unexpanded user queries.

It is also important to note that the WordNet ontology expansion has inverse results when applied to expand the answer terms. This is due to the fact that the WordNet ontology is not suitable for expanding Quranic terms, using WordNet to expand the answer terms adds noise in the metadata. Hence this deteriorates the overall results. Experiments demonstrate that W*ordNet expansion* applied to *Questions only* produces 63% successful results under the Enhanced Parameter Settings (EPS). This success ratio falls to 46% when *WordNet expansion* is applied to *Both (Question & Document)*.

A critical side effect of expanding the answer terms using the WordNet ontology is the amount of processing time. Preliminary experiments were conducted using a PC with decent

configuration (Dual Core with 8GB Ram) but it took over 4.5 hours dedicated processing time, which is not acceptable. The processing time could be improved by involving offline pre-processing of the answer corpus and making use of the metadata instead of processing the answer corpus on the fly and using the metadata residing in the RAM. This simple step could save time but does not contribute any improvement in results. Hence it was not applied during this research. Because of the above observations, in later experiments the WordNet ontology was used for expanding the user query only and not the answer documents.

Another consideration at this stage was the pre-processing of the user questions to identify which question terms should be expanded. A few sub-experiments were conducted to identify the best option. All terms were tagged for their parts of speech using a POS tagger. Firstly, *Noun* and *Verb* tagged words were expanded using the WordNet ontology. Secondly, all question words were processed as *Noun* and *Verb* at the same time. These experiments did not show any significant difference in overall results. Hence all question words were processed using the WordNet ontology.

Lastly, at times WordNet term expansion introduced duplicate words in the question. These duplicate words were ignored as they added an untrue frequency weight for such terms. For reference a sample of WordNet expanded user query has been provided in Appendix 5.

5.1.4 Effect of Cosine Similarity Method

Preliminary experiments conducted to study the effects of similarity methods suggest that the LESK Similarity Method performs better than the Cosine Similarity Method. The LESK Similarity Method produced 50% successful results as compared to the Cosine Similarity Method, which produced 34% successful results when tested under Simple Parameter Settings (SPS).

The above figures remained the same regardless of whether the Snowball or Porter stemmer was used. Similarly different experiments on different term weight functions were conducted to observe the effect. But the trend was maintained across them too. The OKAPI term weighting function produced better results of 58% with the LESK Similarity Measure under Enhanced Parameter Settings (EPS) as compared to the Cosine Similarity Method, which produced 52% correct results.

Based on the above experiments it was identified that the LESK Similarity Method performs relatively better than the Cosine Similarity Method. Hence LESK was used across later detailed experiments.

5.2 Significant Experimental Results

In order to conduct detailed experiments, an additional software program was developed to run experiments in batch mode. This was done to ensure that the required permutations of parameters were tested in full detail. This piece of software facilitates running a group of experiments simultaneously under similar parameters. These experiments have been further sub-grouped at the level of precision rank as discussed in Section 4.4.

The Model Question Set used to test the PARMS Methodology in each experiment includes 100 questions and 156 expected correct answers. The development and design of the PARMS Model QA Set has been discussed in detail in Section 4.1.

It is important to note that in the experimental results presented below, the Mean Reciprocal Rank (MRR) has not been calculated for the P@1 experiments. This is because MRR for every P@1 experiment is implied as 1 in all cases.

5.2.1 SPS Experiment Results

The following table presents the group of experiments that were carried out for precision at different ranks using Simple Parameter Settings (SPS):

(i) Precision at 1 – P@1

The average correct answers for this group of experiments are 2.75 per question. The group of experiments run with SPS for P@1 yields minimum 1 correct answer when tested using the Rashad translation. The following, table 5.1, presents the results of the experiments.

Translator	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP
Maulana	3	3	0.03	0.0300	0.0192	0.0300
Pickthal	4	4	0.04	0.0400	0.0256	0.0400
Yusuf Ali	3	3	0.03	0.0300	0.0192	0.0300
Sher Ali	3	3	0.03	0.0300	0.0192	0.0300
Sarwar	2	2	0.02	0.0200	0.0128	0.0200
Shakir	3	3	0.03	0.0300	0.0192	0.0300
Rashad	1	1	0.01	0.0100	0.0064	0.0100
Khan	2	2	0.02	0.0200	0.0128	0.0200
Hilali & Mohsin	2	2	0.02	0.0200	0.0128	0.0200
Farid-ul-Haq	4	4	0.04	0.0400	0.0256	0.0400
All (Parallel)	2	2	0.02	0.0200	0.0128	0.0200
Parallel-Corpus	4	4	0.04	0.0400	0.0256	0.0400

Table 5.1: SPS Experiment Result for P@1

The accuracy for *Parallel-Corpus* mode was only 0.04. This mode proved useful for experiments using Basic and Enhanced Parameter Settings (discussed in the next experiments) but it did not produce any significant results for the SPS – P@1 group of experiments, because experiments run using Pickthal and Farid-ul-Haq as individual translations also produce the maximum correct answers (4) for this result set.

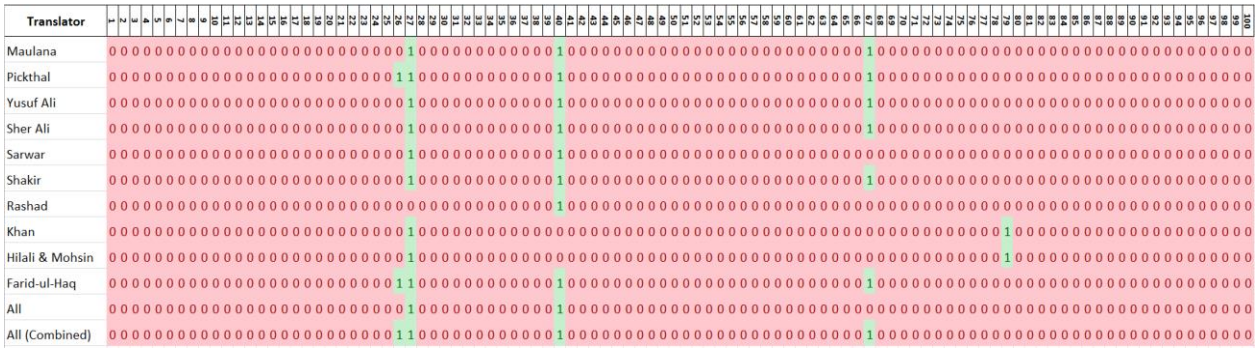


Figure 5.1: SPS Bitmap Image for P@1

As shown in the above bitmap table, the correct answers identified by the PARMS methodology using SPS for P@1 lay among question IDs 26, 27, 40, 67 and 79. The method produced the best results for Question ID#27, as shown in figure 5.1, where 11 out of 12 experiments were able to identify the correct answer.

```
<Question id="27" enabled="true">
  <Text>Do men and women have equal status in Islam in respect of their deeds?</Text>
  <Answers>
    <Answer rank="1">004:124</Answer>
    <Answer rank="2">016:097</Answer>
    <Answer rank="3">033:035</Answer>
    <Answer rank="4">040:040</Answer>
    <Answer rank="5">057:012</Answer>
    <Answer rank="6">057:018</Answer>
  </Answers>
  <Topic>Islamic Ruling:Equality</Topic>
  <SourceContext>
    <Source>http://www.studythequran.com/ShowQandA.php?qa_id=128</Source>
    <DateTaken>2013-01-22</DateTaken>
    <QuestionDate> Unknown </QuestionDate>
    <QuestionType> FAQ </QuestionType>
  </SourceContext>
</Question>
```

Figure 5.2: Model Question and Answer Pair ID# 27

Further analysis of Q&A ID #27 (figure 5.2), shows that this question has six potential correct answers. But it is interesting to note that all successful experiments produce rank 3 033:035 as the correct answer. The Quranic verse for reference number 033:035 is as follows:

> *"Surely the men who submit and the women who submit, and the believing men and the believing women, and the obeying men and the obeying women, and the truthful men and the truthful women, and the patient men and the patient women and the humble men and the humble women, and the almsgiving men and the almsgiving women, and the fasting men and the fasting women, and the men who guard their private parts and the women who guard, and the men who remember Allah much and the women who remember-- Allah has prepared for them forgiveness and a mighty reward."*
> *033:035* *Translated by Shakir*

Simple Parameter Settings (SPS) use the *document frequency count* function to measure the weights of terms in a given document (verse in this case). Close examination shows that verse 033:035 has been a successful selection for a given question due to the unusual frequency of the terms "men" and "women" in this verse.

(ii) Precision at 5 – P@5

The Simple Parameter Settings (SPS) for P@5 produced better than double the result of SPS for P@1 i.e. maximum success score for SPS – P@1 = 0.04 and SPS – P@5 = 0.10. The *All(Parallel)* mode of translation has one of the bottom-line successful results. The *Parallel-Corpus* mode performed better than it did with P@1. Yet it did not produce the best results. Hence the parallel-corpus method incorporated in the PARMS methodology does not contribute a great deal to overall success. Table 5.2 shows these results in detail.

Translator	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP	MRR
Maulana	6	6	0.06	0.0600	0.0385	0.0395	0.6583
Pickthal	10	10	0.10	0.1000	0.0641	0.0573	0.5733
Yusuf Ali	8	8	0.08	0.0800	0.0513	0.0462	0.5771
Sher Ali	7	7	0.07	0.0700	0.0449	0.0425	0.6071
Sarwar	9	9	0.09	0.0900	0.0577	0.0412	0.4574
Shakir	5	5	0.05	0.0500	0.0321	0.0375	0.7500
Rashad	7	7	0.07	0.0700	0.0449	0.0303	0.4333
Khan	9	9	0.09	0.0900	0.0577	0.0445	0.4944
Hilali & Mohsin	9	9	0.09	0.0900	0.0577	0.0445	0.4944
Farid-ul-Haq	5	5	0.05	0.0500	0.0321	0.0433	0.8667
All (Parallel)	5	5	0.05	0.0500	0.0321	0.0293	0.5567
Parallel-Corpus	8	8	0.08	0.0800	0.0513	0.0528	0.6604

Table 5.2: SPS Experiment Result for P@5

The accuracy of the Pickthal translation is best at P@5 = 0.10. The translations of Shakir, Farid-ul-Haq and the *All (Parallel)* translations mode were the worst at P@5 = 5. It is interesting to note that the Pickthal translation was authored in 1930 and uses an 'old' style of English. The Shakir translation is more recent and was first published in 1982, and the

Farid-ul-Haq translation published in 2002, is the latest translation used by the PARMS Methodology. Both these newer translations use more contemporary English. As discussed earlier the model questions have been taken from real Q&A resources. These questions use natural language and have been asked during the modern era (1995 – 2013). Despite this fact, interestingly Pickthal's translation produces double the successful results of Shakir and Farid-ul-Haq translations.

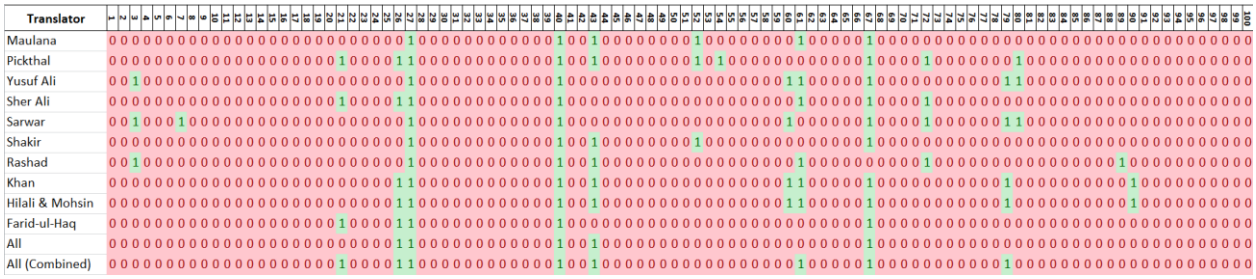


Figure 5.3: SPS Bitmap Image for P@5

The bitmap table in figure 5.3 indicates the success of Question ID# 27 and 40 with every translation mode. It can be observed from this figure that there is more diversity in successful questions for this set of experiments than for SPS – P@1, which involved only 5 questions.

(iii) Precision at 10 – P@10

The average success of correct answers with Simple Parameter Settings (SPS) for P@10 is 11.25 for each question. The experiment carried out with the *All (Parallel)* translations mode produced the worst results for this group of experiments. Table 5.3 shows the results:

Translator	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP	MRR
Maulana	9	9	0.09	0.0900	0.0577	0.0438	0.4872
Pickthal	13	14	0.13	0.1300	0.0897	0.0605	0.4734
Yusuf Ali	11	11	0.11	0.1100	0.0705	0.0500	0.4541
Sher Ali	11	11	0.11	0.1100	0.0705	0.0483	0.4389
Sarwar	12	13	0.12	0.1200	0.0833	0.0450	0.3726
Shakir	12	12	0.12	0.1200	0.0769	0.0467	0.3890
Rashad	10	10	0.10	0.1000	0.0641	0.0341	0.3411
Khan	14	14	0.14	0.1400	0.0897	0.0506	0.3617
Hilali & Mohsin	14	14	0.14	0.1400	0.0897	0.0508	0.3630
Farid-ul-Haq	8	8	0.08	0.0800	0.0513	0.0471	0.5890
All (Parallel)	6	6	0.06	0.0600	0.0385	0.0293	0.4824
Parallel-Corpus	15	15	0.15	0.1500	0.0962	0.0619	0.4126

Table 5.3: SPS Experiment Result for P@10

The Parallel-Corpus mode on the other hand produced the best results for this experiment group with 0.15 accuracy in retrieving correct answers. The bitmap table in figure 5.4 shows the different degrees of different success of the translations. It is noteable that SPS for P@10 produces success for a few common questions, i.e. QID# 27, 40 and 67.

Broadening the precision level to 10 has allowed the Rashad tranlation to produce the correct result for QID# 27 (figure 5.2), which it was not able to identify for P@1. This is because the potential correct answer for P@1 in fact occurs at P@2, which becomes evident when P@10 is used.

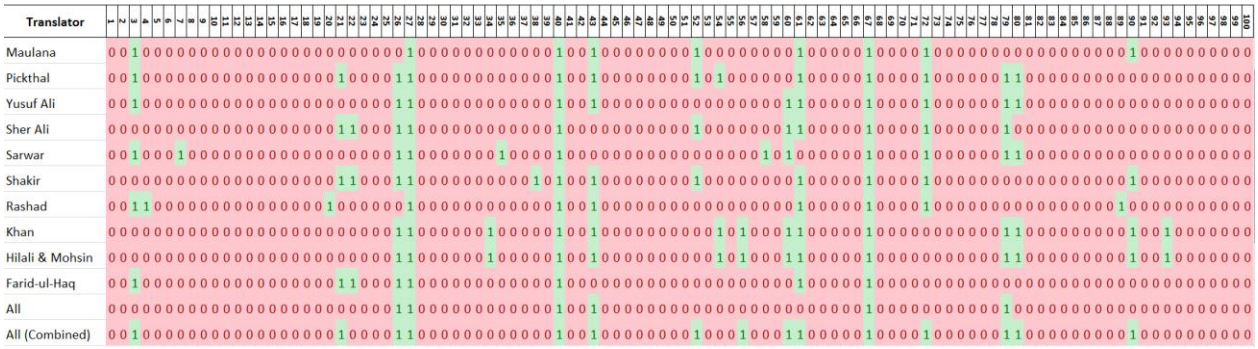


Figure 5.4: SPS Bitmap Image for P@10

The above experiment groups conducted for P@1, P@5 and P@10 suggest that the statistical method based on frequency of terms alone might not be very useful for the Quranic dataset. This fact is evident from the results suggested in the tables in these sections.

5.2.2 BPS Experiment Results

Experiments conducted under Basic Parameters Settings (BPS) for precision at different ranks show improved and interesting results. As compared to SPS, these parameter settings include an element popular in NLP techniques, i.e. they use stopwords removal and stemming of words along with advanced statistical methods of term frequency weighting.

(i) Precision at 1 – P@1

The PARMS Methodology is unable to retrieve answers for all model questions for the group of experiments conducted for BPS – P@1. The system is unable to provide any answer for the following 2 model questions; QID#01 and QID#87. The answers to these questions are not found despite using seven out of ten author translations. *All (Parallel)* and *Parallel-Corpus* translation modes are able to identify the answer for QID#01 but not able to produce any results for QID#87. This peculiar case was not identified during experiments conducted under SPS. On closer analysis it was observed that this happens because BPS

removes stopwords whereas SPS does not. For illustration, the example of QID#87 is presented.

```
<Question id="87" enabled="true">
  <Text>How near is Allah to us?</Text>
  <Answers>
    <Answer rank="1">050:016</Answer>
  </Answers>
  <Topic>Faith:Belief</Topic>
```

Figure 5.5: Model Question QID#87

The following image shows the Question Terms produced for QID#01 when processed using Simple Parameter Settings (SPS). As SPS does not take away the stopwords, all words in the question are regarded as question terms.

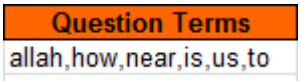


Figure 5.6: Question Terms for QID#87 using SPS

It can be observed that all words in QID#87 are 'less informative' terms for the Quranic dataset. As a result these words return a huge number of potential answers due to the fact that these words exist frequently in many Quranic verses. The following figure predicts that the correct answer is located at the 3,748th location within 4,274 potential answers identified from 6,236 a total of Quranic verses.

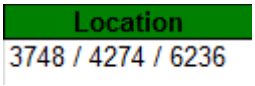


Figure 5.7: Location of Correct Answer for QID#87 using SPS

Stopwords removal takes away the less informative words. Hence when QID#87 is processed using Basic Parameter Settings (BPS) leaves the Question Terms set is empty. This can be observed from the following table 5.4 and bitmap image presented in figure 5.8:

Translator	Total Retrieved Results	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP
Maulana	98	17	17	0.17	0.1735	0.1090	0.1700
Pickthal	98	9	9	0.09	0.0918	0.0577	0.0900
Yusuf Ali	98	13	13	0.13	0.1327	0.0833	0.1300
Sher Ali	98	10	10	0.10	0.1020	0.0641	0.1000
Sarwar	98	22	22	0.22	0.2245	0.1410	0.2200
Shakir	98	17	17	0.17	0.1735	0.1090	0.1700
Rashad	98	15	15	0.15	0.1531	0.0962	0.1500
Khan	99	15	15	0.15	0.1515	0.0962	0.1500
Hilali & Mohsin	99	15	15	0.15	0.1515	0.0962	0.1500
Farid-ul-Haq	99	22	22	0.22	0.2222	0.1410	0.2200
All (Parallel)	99	26	26	0.26	0.2626	0.1667	0.2600
Parallel-Corpus	99	25	25	0.25	0.2525	0.1603	0.2500

Table 5.4: BPS Experiment Result for P@1

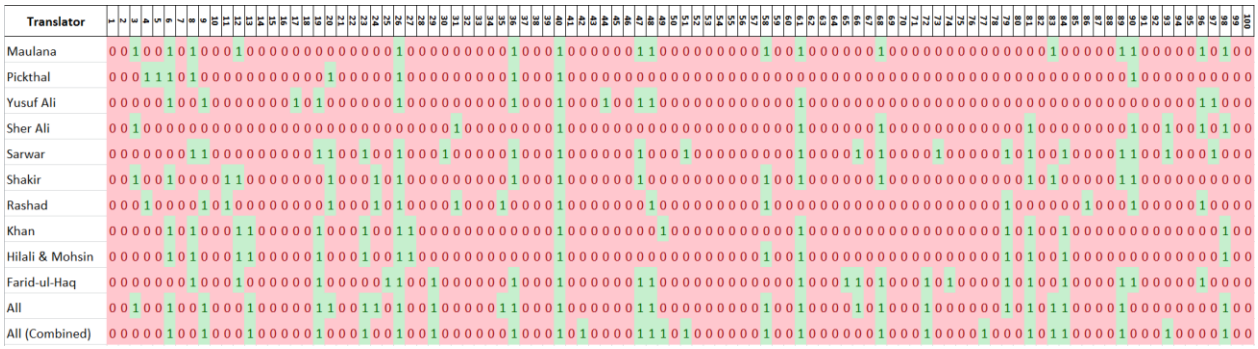


Figure 5.8: BPS Bitmap Image for P@1

(ii) Precision at 5 – P@5

The experiments conducted under BPS for precision at the top 5 ranks suggest a gradual improvement in results, but this does not ensure an answer to all questions. As indicated earlier, BPS – P@5 does not ensure an answer to all questions. The following table presents detailed results of the experiments:

Translator	Total Retrieved Results	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP	MRR
Maulana	98	29	29	0.29	0.2959	0.1859	0.2012	0.6937
Pickthal	98	20	21	0.20	0.2041	0.1346	0.1255	0.6400
Yusuf Ali	98	37	42	0.37	0.3776	0.2692	0.2157	0.5838
Sher Ali	98	33	35	0.33	0.3367	0.2244	0.1860	0.5586
Sarwar	98	38	40	0.38	0.3878	0.2564	0.2763	0.7263
Shakir	98	31	32	0.31	0.3163	0.2051	0.2125	0.6828
Rashad	98	31	32	0.31	0.3163	0.2051	0.2160	0.6941
Khan	99	33	36	0.33	0.3333	0.2308	0.2102	0.6444
Hilali & Mohsin	99	32	35	0.32	0.3232	0.2244	0.2113	0.6698
Farid-ul-Haq	99	33	36	0.33	0.3333	0.2308	0.2518	0.7798
All (Parallel)	99	41	44	0.41	0.4141	0.2821	0.3098	0.7598
Parallel-Corpus	99	49	55	0.49	0.4949	0.3526	0.3235	0.6748

Table 5.5: BPS Experiment Result for P@5

The Parallel-Corpus shows the best performance for most evaluation metrics. The following bitmap image (figure 5.9), shows more diversity in successful questions than SPS for P@5.

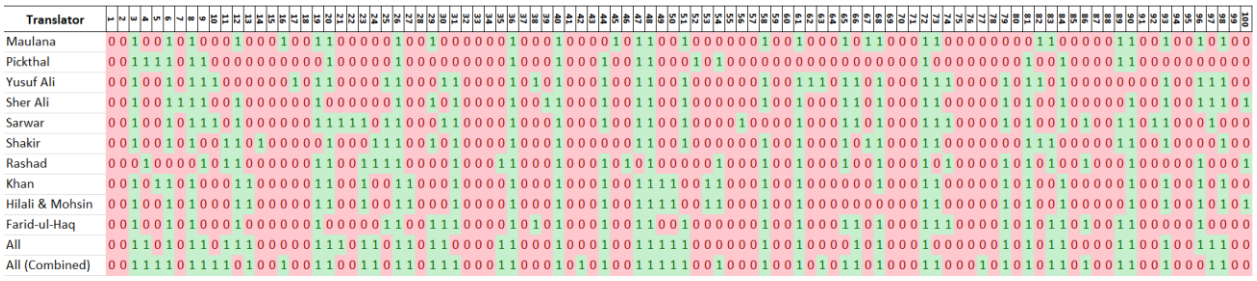


Figure 5.9: BPS Bitmap Image for P@5

(iii) Precision at 10 – P@10

The following table shows that Basic Parameter Settings (BPS) are not able to produce answers for all model questions for precision at 10 also. This is due to the same reason as discussed earlier. It is notable from the results below that BPS – P@10 allows a broader answer window; hence more correct answers are retrieved. The results show the better performance of the Parallel-corpus on Single and All (Parallel) translation modes.

Translator	Total Retrieved Results	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP	MRR
Maulana	98	34	37	0.34	0.3469	0.2372	0.2085	0.6148
Pickthal	98	36	41	0.36	0.3673	0.2628	0.1484	0.4147
Yusuf Ali	98	42	50	0.42	0.4286	0.3205	0.2194	0.5315
Sher Ali	98	45	49	0.45	0.4592	0.3141	0.2045	0.4498
Sarwar	98	43	48	0.43	0.4388	0.3077	0.2765	0.6582
Shakir	98	34	36	0.34	0.3469	0.2308	0.2169	0.6337
Rashad	98	37	39	0.37	0.3776	0.2500	0.2197	0.6012
Khan	99	45	51	0.45	0.4545	0.3269	0.2244	0.5058
Hilali & Mohsin	99	44	50	0.44	0.4444	0.3205	0.2263	0.5220
Farid-ul-Haq	99	43	48	0.43	0.4343	0.3077	0.2601	0.6292
All (Parallel)	99	51	59	0.51	0.5152	0.3782	0.3092	0.6359
Parallel-Corpus	99	57	66	0.57	0.5758	0.4231	0.3343	0.5984

Table 5.6: BPS Experiment Result for P@10

Similarly, the better performance trend for Parallel-Corpus is also apparent across other evaluation metrics. An important observation is that the Mean Reciprocal Rank (MRR) for all experiments ranges between fourth and sixth rank. This indicates that it is more usual to find the first correct answer to each question towards the middle. It also means that perhaps after BPS – P@6 chances to find a successful answer decrease.

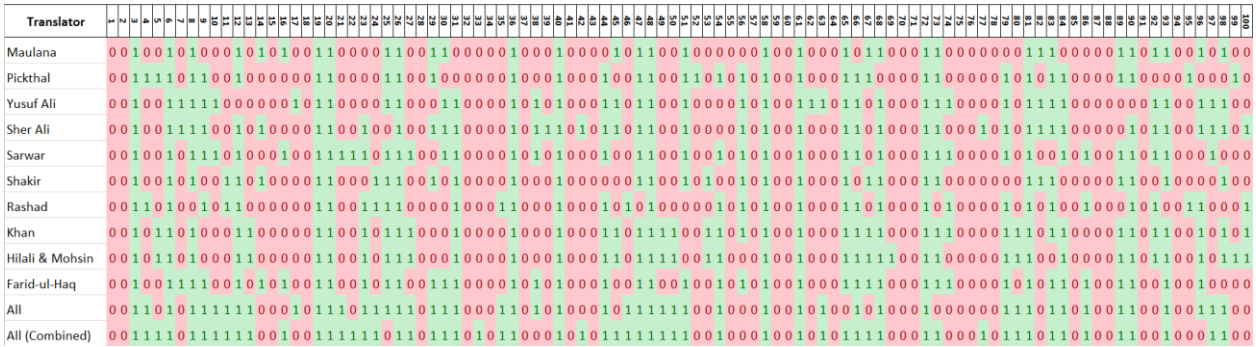


Figure 5.10: BPS Bitmap Image for P@10

The above bitmap image (figure 5.10), shows the diversity of successful answers. It can be noted that there are some questions for which all translation modes are able to produce correct results, e.g. QID#3, QID#6, QID#19, QID#20, and QID#27, QID#40. This could be

due to the fact that the answer verses for these questions hold more informative terms related to the concepts asked in the model questions.

(iv) Overall Success Trends for BPS

The variation in success trends for experiments under Basic Parameter Settings (BPS) at different precisions can be observed from the graph shown in figure 5.11. There is a general trend in success rate for single translations. A visible improvement in success is observed with the shift of precision from 1 to 5 with the maximum difference being in Yusuf Ali's translation. It can also be observed that there is less improvement in successful results when the answer window is expanded to P@10.

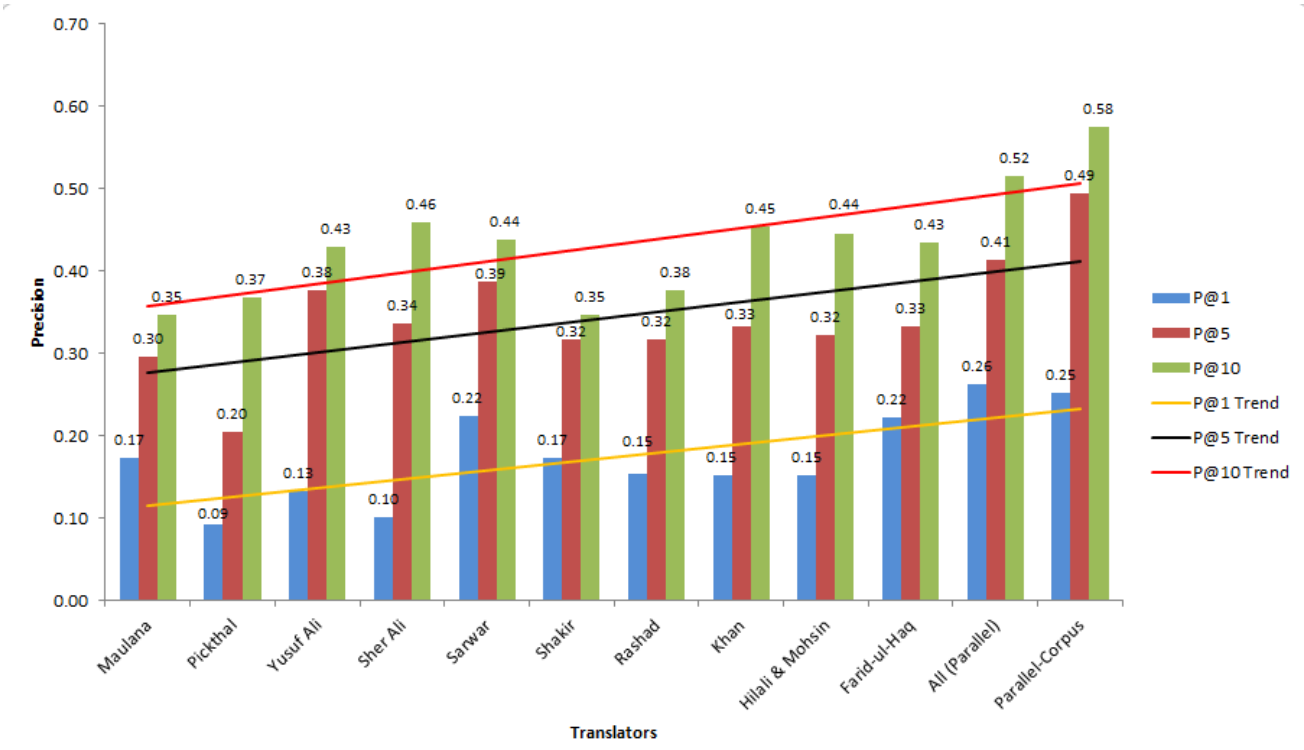


Figure 5.11: Graph of Overall Success Trends for BPS Experiments Result

The All (Parallel) translation mode performed slightly better than the Parallel-Corpus mode for BPS – P@1 but this trend changed for drastically for BPS – P@5 and was maintained for experiments conducted under BPS – P@10. It is also notable that the overall performance gradient has remained steady across all precision ranks but there is variation in performance in terms of correct results.

5.2.3 EPS Experiment Results

(i) Precision at 1 – P@1

Experiments conducted under Enhanced Parameter Settings (EPS) produce results for all 100 Model Questions unlike BPS. These experiments produce a maximum 29 correct results having a precision value of 0.29 for P@1. The correct retrieved results and answers are the same in the case of P@1 due to the answer window (result set) size being 1. This will also lead to the same values for Accuracy, Precision and Mean Average Precision. The following table shows these results:

Translator	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP
Maulana	21	21	0.21	0.2100	0.1346	0.2100
Pickthal	13	13	0.13	0.1300	0.0833	0.1300
Yusuf Ali	19	19	0.19	0.1900	0.1218	0.1900
Sher Ali	18	18	0.18	0.1800	0.1154	0.1800
Sarwar	21	21	0.21	0.2100	0.1346	0.2100
Shakir	21	21	0.21	0.2100	0.1346	0.2100
Rashad	20	20	0.20	0.2000	0.1282	0.2000
Khan	17	17	0.17	0.1700	0.1090	0.1700
Hilali & Mohsin	18	18	0.18	0.1800	0.1154	0.1800
Farid-ul-Haq	19	19	0.19	0.1900	0.1218	0.1900
All (Parallel)	26	26	0.26	0.2600	0.1667	0.2600
Parallel-Corpus	29	29	0.29	0.2900	0.1859	0.2900

Table 5.7: EPS Experiment Result for P@1

It can be observed from the bitmap image in figure 5.12 that QID#87 produces some results under EPS. This is merely due to WordNet expansion of existing terms before stopwords removal. But it is also notable that no experiment produced the correct answer for this question because the existing words were all insignificant words, and hence added no value to the answer search.

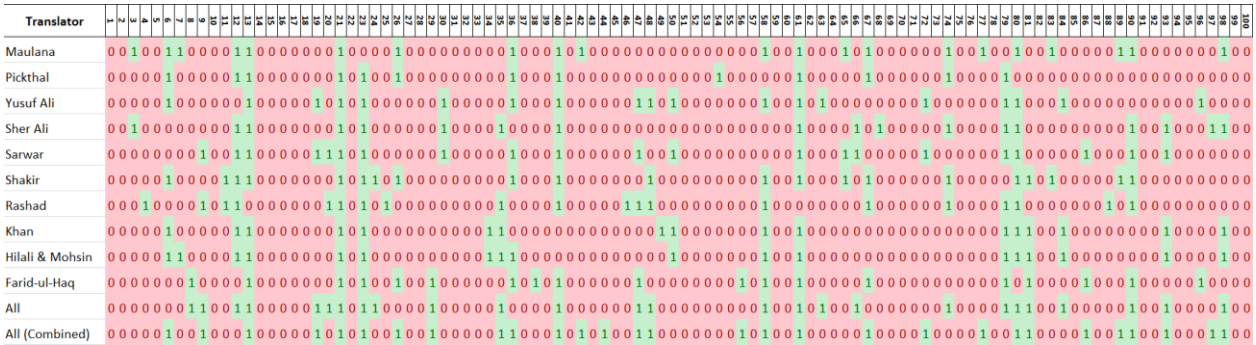


Figure 5.12: EPS Bitmap Image for P@1

(iii) Precision at 5 –P@5

The success rate improves with an increase in answer window size. This can be observed from table 5.8 where the minimum number of correct answers is 34 while the maximum stand at 41 for individual translations. There is a significant improvement in the results by using All(Parallel) corpus. This further improves to 53 when using the Parallel-Corpus. The number of correct answers also increases from 46 to 58. So both precision and recall is improved by using the Parallel-Corpus. Mean Average Precision follows the same trend. Details of the results are shown in table 5.8:

Translator	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP	MRR
Maulana	41	46	0.41	0.4100	0.2949	0.2845	0.7020
Pickthal	34	38	0.34	0.3400	0.2436	0.2177	0.6353
Yusuf Ali	39	45	0.39	0.3900	0.2885	0.2601	0.6786
Sher Ali	37	42	0.37	0.3700	0.2692	0.2519	0.6788
Sarwar	38	40	0.38	0.3800	0.2564	0.2737	0.7180
Shakir	39	43	0.39	0.3900	0.2756	0.2798	0.7197
Rashad	40	41	0.40	0.4000	0.2628	0.2708	0.6750
Khan	39	47	0.39	0.3900	0.3013	0.2467	0.6380
Hilali & Mohsin	38	46	0.38	0.3800	0.2949	0.2510	0.6671
Farid-ul-Haq	40	43	0.40	0.4000	0.2756	0.2757	0.6850
All (Parallel)	47	52	0.47	0.4700	0.3333	0.3319	0.7092
Parallel-Corpus	53	58	0.53	0.5300	0.3718	0.3815	0.7182

Table 5.8: EPS Experiment Result for P@5

It is important to note that the MRR value is approximately the same for both (some of) the translators and the parallel corpus. So the use of the parallel corpus does not make any significant improvement in finding the answer on a particular rank.

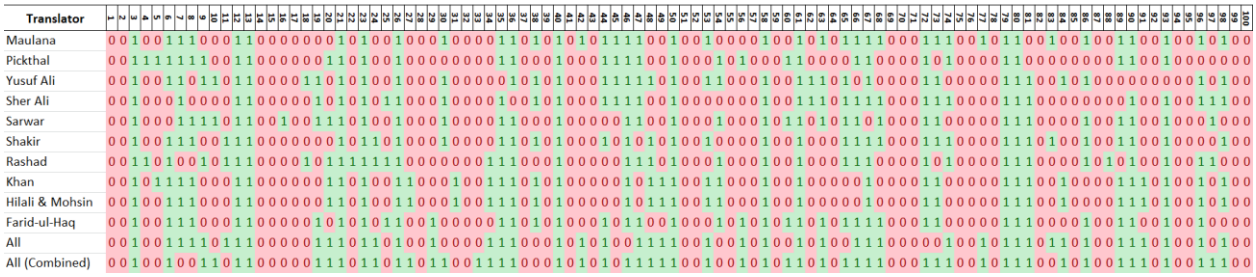


Figure 5.13: EPS Bitmap Table for P@5

The bitmap table in figure 5.13 for EPS – P@5 shows more regions coloured in green (indicated as '1'). This indicates that for certain questions almost all translation modes have successfully identified correct answers.

(iii) Precision at 10 –P@10

The PARMS Methodology generates the best results under the Enhanced Parameter Settings (EPS) with answer window size P@10. The best results are achieved by the Parallel-Corpus translation mode. This can be observed from the following table:

Translator	Correct Retrieved Results	Correct Retrieved Answers	Accuracy	Precision	Recall	MAP	MRR
Maulana	48	55	0.48	0.4800	0.3526	0.2946	0.6187
Pickthal	41	47	0.41	0.4100	0.3013	0.2252	0.5475
Yusuf Ali	48	56	0.48	0.4800	0.3590	0.2696	0.5750
Sher Ali	46	51	0.46	0.4600	0.3269	0.2632	0.5706
Sarwar	49	52	0.49	0.4900	0.3333	0.2895	0.5879
Shakir	50	57	0.50	0.5000	0.3654	0.2964	0.5940
Rashad	47	51	0.47	0.4700	0.3269	0.2730	0.5936
Khan	48	57	0.48	0.4800	0.3654	0.2591	0.5436
Hilali & Mohsin	48	57	0.48	0.4800	0.3654	0.2646	0.5559
Farid-ul-Haq	46	49	0.46	0.4600	0.3141	0.2827	0.6110
All (Parallel)	58	65	0.58	0.5800	0.4167	0.3400	0.5994
Parallel-Corpus	63	70	0.63	0.6300	0.4487	0.3970	0.6277

Table 5.9: EPS Experiment Result for P@10

The Maximum correct answers identified are 63/100. This produces the maximum Accuracy and Precision of the method, i.e. 0.63. It is notable that the tendency to find the first correct answer somewhere in the upper half of the result set also prevails. This finding is also obvious from the graph in figure 5.15. The graph is indicative of the fact that increasing the size of the result set window improves the probability of finding a correct answer. This improvement trend continues up to the middle of the result set, i.e. up to rank 5-6. After this the results remain steady.

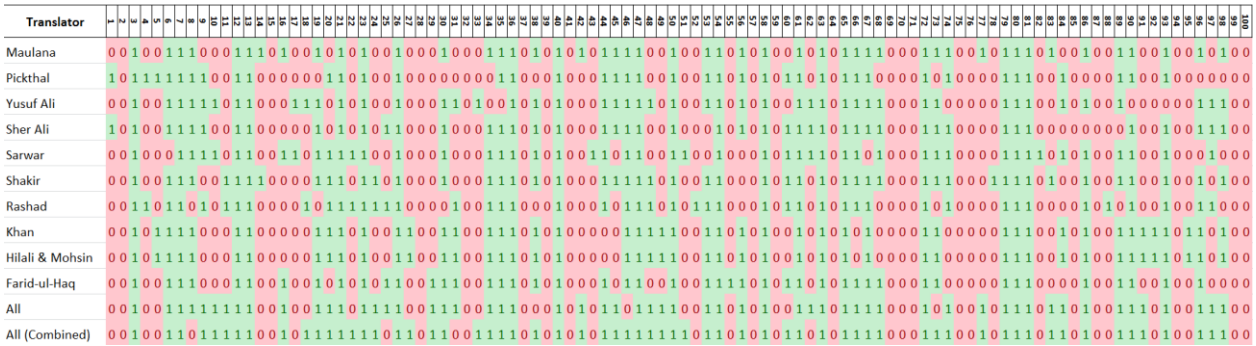


Figure 5.14: EPS Bitmap Table for P@10

The bitmap table shown in figure 5.14 also reflects the best achieved correct answer set for each question against each mode of translation.

(iv) Performance Trend for Parallel-Corpus Mode

The detailed experiments reported in previous sections reveal that the Parallel Corpus mode proposed by the PARMS Methodology produces better results than the use of individual authors or the All (Parallel) mode of translations. The graph shown in figure 5.15 also depicts the fact that a window size of 5-6 contributes the most to the correct answer. The performance of the methodology becomes steady beyond this window size.

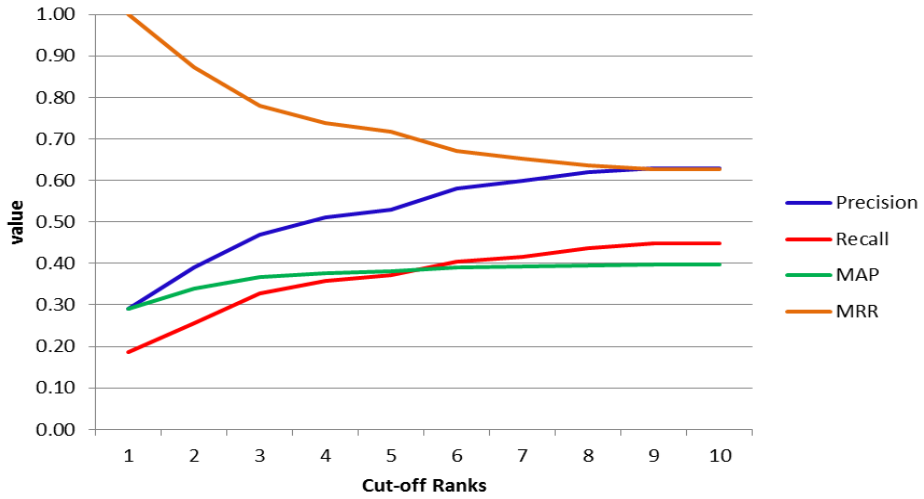


Figure 5.15: Graph of Performance Trends for Parallel-Corpus Mode

(v) Recall by Rank

The 'Recall by Rank' graph below displays the ratios of the answers retrieved at ranks from 1 to 10. The graph (figure 5.16), shows that, in all, approximately 45% of answers are retrieved correctly. Out of these 45% correctly retrieved answers, 18.5% are ranked at position 1.

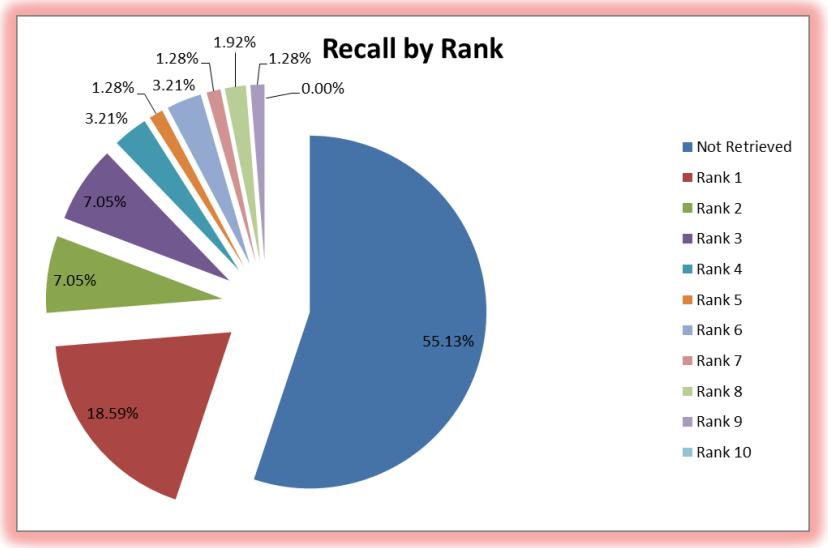


Figure 5.16: Graph of Recall by Rank

Thus 41.4% of correctly retrieved answers are retrieved at rank 1. If we look at the top 3 ranks then this ratio is 32.69% of the total correct answers, making 72.8% of the correctly retrieved answers. This indicates that the system tends to find most of the correct answers towards the higher ranks. In fact 90% of correctly retrieved answers are found at ranks 1-6, while answers found at ranks 7-10 are only 10%. It is important to note that none of correctly retrieved answers is found at rank 10. The probability of finding a correct answer at lower ranks is not only low but it also shows a steadily decreasing trend.

5.3 Evaluation Results Analysis

This section presents the analysis of findings of detailed experiments carried out to test the PARMS Methodology implemented on the PARMS QA Application and applied to the Parallel Corpus of English translations of the Quran. It reflects on how certain novel features presented by the PARMS Methodology contribute to producing better results for the dataset using Model Questions.

5.3.1 Role of Parallel-Corpus

The PARMS Evaluation System facilitates the testing of the PARMS methodology using single-author translation or parallel-translations at once. It is observed from the detailed experiments that use of the All (Parallel) translation mode has the edge over using the single-author translation mode. But the results are best with the Parallel-Corpus mode of translation. The graph presented in figure 5.17 also reflects on how the Parallel-Corpus contributes towards overall successful results.

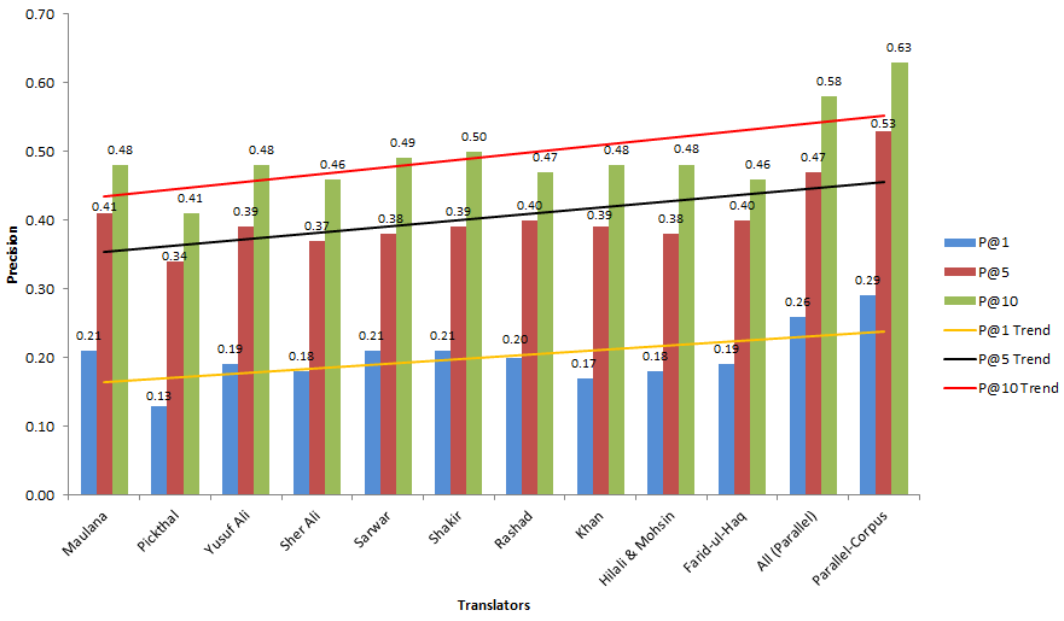


Figure 5.17: Graph of Performance Trend for Parallel-Corpus Mode for EPS Experiments

It can be observed from the graph that the Parallel-Corpus mode produced best results amongst available translation modes regardless of answer-window size.

5.3.2 Role of Multi-Stream Semantic Knowledge

The detailed experiments presented in the previous section support the following hypothesis: multiple-stream semantic knowledge proves helpful in identifying more accurate answers for a natural language question. This hypothesis is true for semantic knowledge embedded using QurAna and QurTerms. The resolution of pronouns found in a verse assists in determining the correct answer. Similarly translating terms into natural language terminology using QurTerms improves understanding of domain-specific terms. But the hypothesis does not hold for semantic knowledge presented in QurSim, which is used for re-ranking.

I – Experiment One

This can be witnessed from the following an experiment where PARMS Model Question ID Number 38 (QID#38) is analysed closely using an early version of the PARMS QA System. The content of the model QID#38 and the correct answer verse as referenced by scholars in answering this question are given below:

QID#38: "Is making a Will recommended by the Quran? How many witnesses are required to testify to the Will? What is the procedure for obtaining witnesses when making the will is necessitated during journey?"

> ***Correct Answer Verse:*** *"'O believers! The evidence among yourselves when death approaches any of you at the time of making a will is of two jut persons from among you or two of the others, when you be on journey in the land, and the calamity of death befall you, Detain them both after prayer, they should swear by Allah if you are in any doubt, that we shall not purchase any wealth in exchange of the oath, even though he may be a near relation, and will not hide the testimony of Allah if we do so, we are surely among the sinners. "*
>
> *005:106* *Translated By Farid-ul-Haq*

The PARMS Evaluation System produces incorrect results for QID#38 when tested against parameter settings indicated in the screen shot presented in figure 5.18. The experiment has been run on Farid-ul-Haq's translation. Less informative words have been eliminated from the natural language question and key terms have been stemmed to derive the root words for the morphological variants of these terms. The IDF term-weighting function has been used for statistical weighting. Terms have been expanded using the WordNet (Miller, Beckwith et al. 1990, Miller 1995, Fellbaum 1998, Pedersen, Patwardhan et al. 2004) ontology.

The results of this experiment propose 002:196 as the top-ranked answer and 065:001 as the second-ranked answer, based on the above-mentioned parameters. Both of these answers are incorrect. The screen shot in figure 5.18 presents these results. The same experiment when run with Pronoun Resolution enabled identifies the anticipated correct answer verse reference. Screen shot 5.19 shows this correct result.

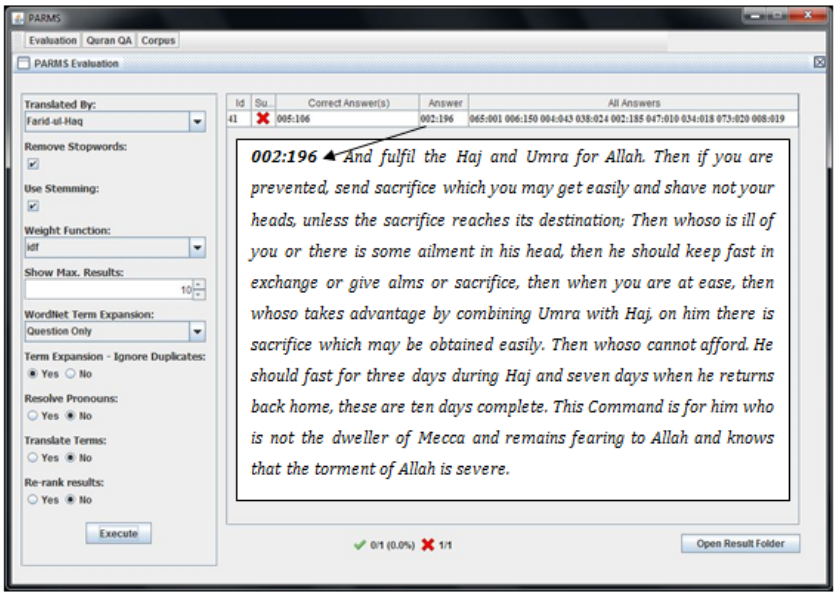


Figure 5.18: Screen Shot of PARMS Results showing incorrect results for QID#38

The results in figure 5.19 show the correct answer verse 005:106 as the top-ranked answer verse whereas 002:196 and 065:001 have been lowered down a rank each. A comparative analysis of both test scenarios show that additional semantic knowledge embedded in the system enables correct answers to move several ranks up. In the given example the correct answer 005:106 was at rank number 13 in the first experiment results. Pronoun resolution

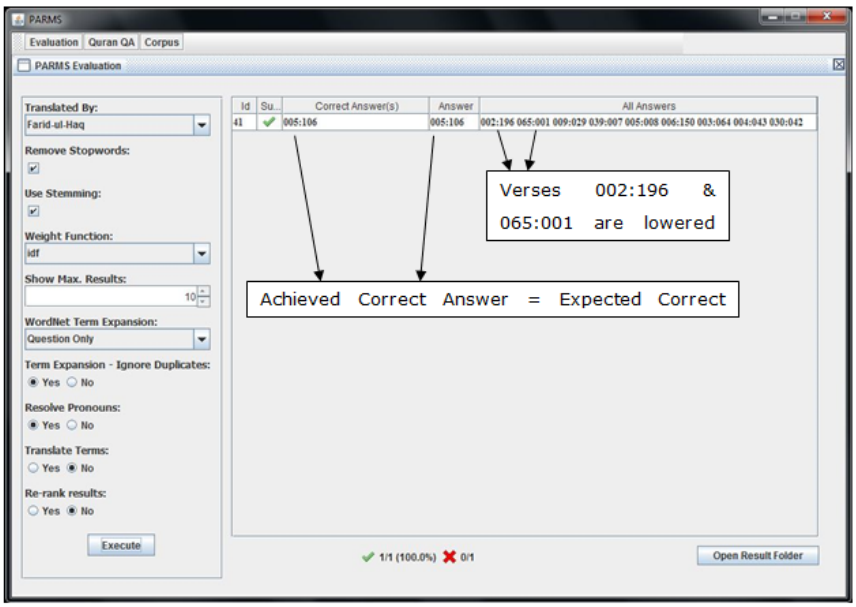


Figure 5.19: Screen Shot of PARMS Results showing correct results for QID#38

of the verse pronouns add to the term weights which in return adds into the overall ranking score of the verse. Term weight calculated for Verse 002:196 is 21.212 whereas the term weight for the top-ranked verse was 29.548, which makes it a potential correct answer.

II – Experiment Two

The role of additional semantic knowledge was studied through several experiments. For example, the following experiment investigates these effects. This experiment was set up in 3 successive stages for detailed observation. The initial test uses Shakir's translation. Stopwords are taken away and stemming is performed using the Snowball stemmer. This part of the experiment uses the *idf* function for term weighting and the LESK method for similarity measurement. The terms have been expanded at the initial level using the WordNet ontology. These can be observed in the screen shot shown in figure 5.20:

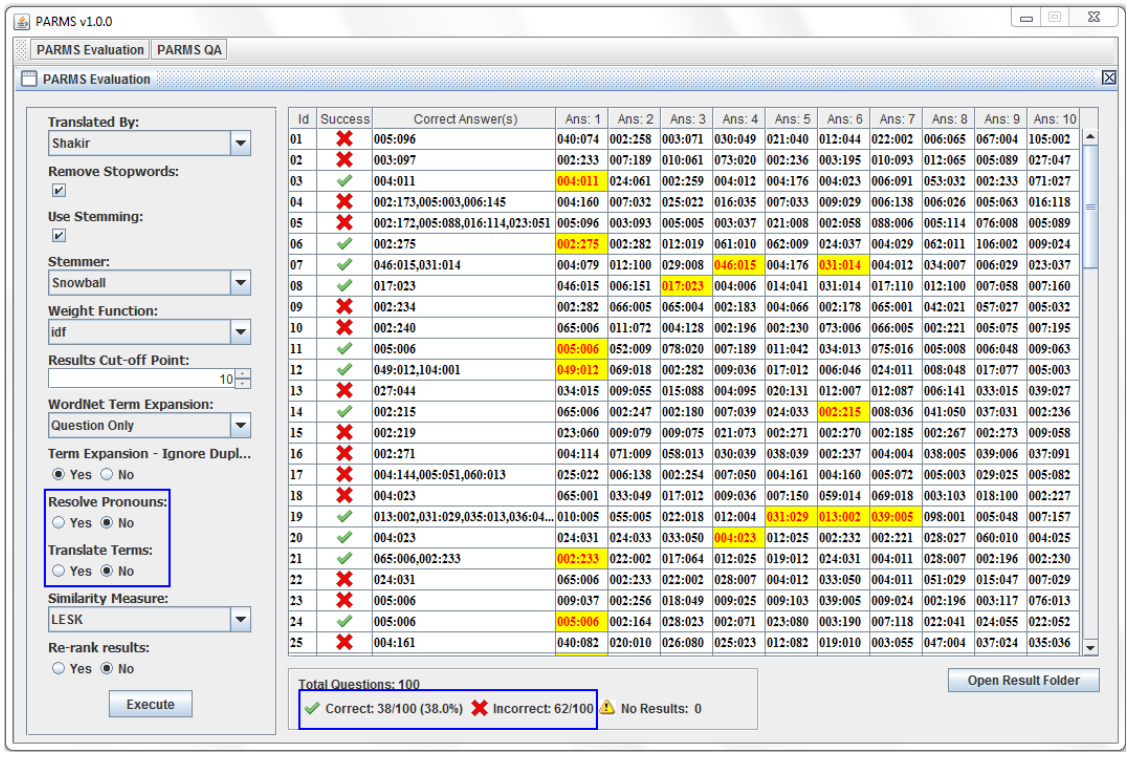


Figure 5.20: Screen Shot of PARMS Evaluation System Result for Experiment Two Stage 1

At the first stage the experiment does not use either the QurAna ontology for pronoun resolution or the QurTerms ontology for translating terms. As indicated, the PARMS Algorithm identified 38 answers successfully and was unable to provide correct answers for 62 model questions. At the second stage of the experiment, additional semantic knowledge was introduced along with the existing parameter setting. The QurAna ontology was used to replace the verse pronouns with their referents. This improved the success of the PARMS

Evaluation System, such that the PARMS Algorithm produced 44 successful results. This is shown in figure 5.21.

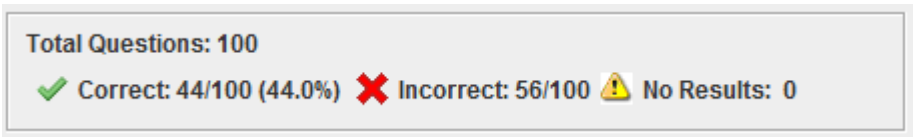


Figure 5.21: PARMS Evaluation System Result for Experiment Two Stage 2

The change Chart (figure 5.22) shows the effect of this additional knowledge on PARMS Model Questions. It is evident that the PARMS Algorithm produces correct results for additional PARMS Model Questions at Stage 2 as compared to results at Stage 1. The following are these additional successful results i.e. QID#54, QID#60, QID#63, QID#66, QID#78, QID#79, QID#80 and QID#86.

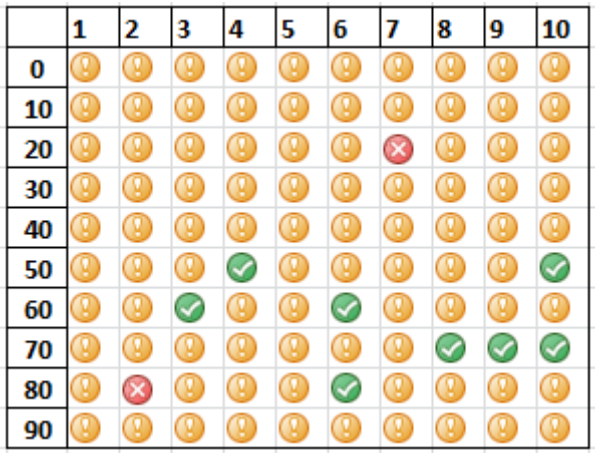


Figure 5.22: Change Chart for Experiment Two Stage 1 & 2

It is observed that alongside the positive impact of additional knowledge, at times there can be negative impacts of this added knowledge on the cumulative results. This can be witnessed in change chart shown in figure 5.22. PARMS Model Questions QID#27 and QID#82 reflect this negative influence. Closer analysis of QID#82 shows that pronoun resolution of the related verses introduced 'noise' in the verse.

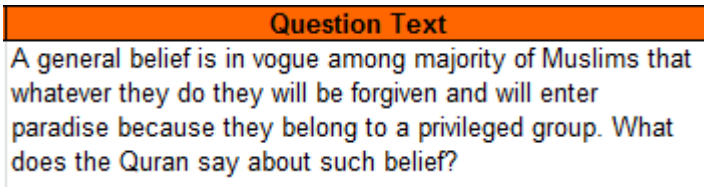


Figure 5.23: PARMS Model Question QID#82

The anticipated correct answer as referenced by scholars for QID#82 is Verse Reference Number 007:169. The *answer window* of 10 has been used for all stages of this experiment.

Location	Answer Text
9 / 1313 / 6236	Then there came after them an evil posterity who inherited the Book, taking only the frail good of this low life and saying: It will be forgiven us. And if the like good came to them, they would take it (too). Was not a promise taken from them in the Book that they would not speak anything about Allah but the truth, and they have read what is in it; and the abode of the hereafter is better for those who guard (against evil). Do you not then understand?

Figure 5.24: Correct Answer 007:169 for PARMS Model Question QID#82

This means if the anticipated answer is found in top ten ranks, it will be considered as the correct answer.

The location of the correct answer as found at stage 1 is 9 whereas 1313 verses were initially identified as potential answers. The cumulative weight of this answer is 15.83. Figure 5.24 shows the answer text along with its location.

The Quranic verse shown above makes use of several pronouns. It is not possible to determine the true meaning of the verse unless these pronouns are resolved to their referents. At stage 2 of this experiment this additional semantic knowledge for this verse is added using the QurAna ontology. The related referents of the pronouns used in this Quranic verse are given below.

"Children,of,Israel,Jew,Rabbis,who,do,not,follow,Torah,Jew,Rabbis,who,do,not,follow,Torah,Jew,Rabbis,who,do,not,follow,Torah,Jew,Rabbis,who,do,not,follow,Torah,Jew,Rabbis,who,do,not,follow,Torah,various,kinds,of,forbidden,sins,Jew,Rabbis,who,do,not,follow,Torah,various,kinds,of,forbidden,sins,Jew,Rabbis,who,do,not,follow,Torah,Jew,Rabbis,who,do,not,follow,Torah,Jew,Rabbis,who,do,not,follow,Torah,the,Torah,(Muttaqun),the,pious,,the,righteous,,God,fearing,Jew,Rabbis,who,do,not,follow,Torah"

Pronouns resolution of this verse deteriorates the term weight of this verse. The new location of this verse falls at 20th rank out of 1586 total potential answers. The pronoun referent terms added to the verse in fact introduce noise in this case. At the third stage of the experiment semantic knowledge is further enriched using the QurTerms ontology. This additional knowledge improves the performance of the PARMS Algorithm bringing the successful results up to 50%, as shown in figure 5.25.

Total Questions: 100

Correct: 50/100 (50.0%) Incorrect: 50/100 No Results: 0

Figure 5.25: PARMS Evaluation System Result for Experiment Two Stage 3

The PARMS Methodology enriched by the QurTerms ontology improved the results of the experiment. The evaluation system successfully identified answers to six new PARMS Model Questions, namely QID#13, QID#23, QID#30, QID#46, QID#48 and QID#53.

The following Change Chart shown (figure 5.26), depicts the change between stage 1 and stage 2 of the experiment. From this detailed experiment discussion, it can be concluded that the PARMS Algorithm performance improves gradually as more knowledge is embedded within the PARMS Methodology.

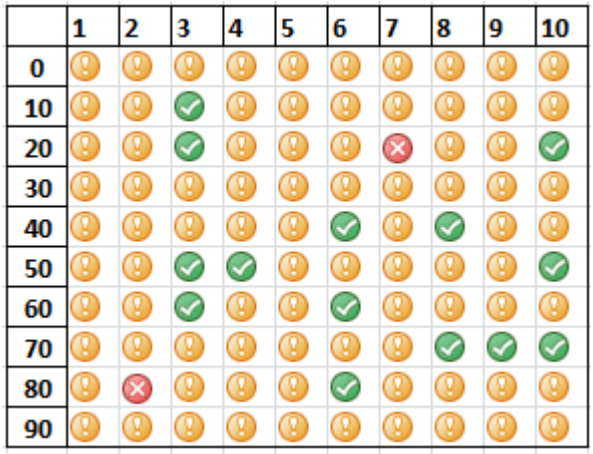


Figure 5.26: Change Chart for Experiment Two Stage 1 & 3

III – Experiment Three

A shift in the favourable results of above discussed hypothesis was noticed when the QurSim ontology was used to re-rank answers produced by the PARMS Methodology. The assumption behind re-ranking was that as QurSim presents the degree of similarity between Quranic verses, it should be possible to use this additional knowledge to boost the ranking of answers based on the fact that a similar-answer weight can be added to the weight of a selected answer.

The implemented algorithm was enabled to sum up weights of similar verses that occur in the set of potential answers such that a verse at a lower rank volunteers its weight to similar ones at higher ranks. This should ideally enhance the chances of the verses discussing the core concepts asked in the question, to move towards the top and be included as an answer. But in practice this additional knowledge invalidly boosts some answers that are not relevant. As a result they replace some existing correct answers pushing the correct ones lower down in rank. This generally decreases the number of correct answers. For example the best results of the PARMS Methodology produced under Enhanced Parameter Settings (EPS) using a result window of P@10 is 63%. This result declines to only 35% when the QurSim ontology is used for re-ranking the result set. This

change of results between experiments can be observed in the Change Chart Diagrams shown in figure 5.27. The diagram at the left is the PARMS Best Results, i.e. 63%, without re-ranking, whereas the right-hand diagram shows the re-ranked results i.e. 35% correct answers.

Despite the fact that re-ranking does not ensure potential answers, in most cases the correct answer is moved higher in rank when processed using the QurSim ontology although it does not necessarily fall in the correct answer window. The following example shows the working of PARMS Model Question QID#04. This can have following 3 possible correct answers.

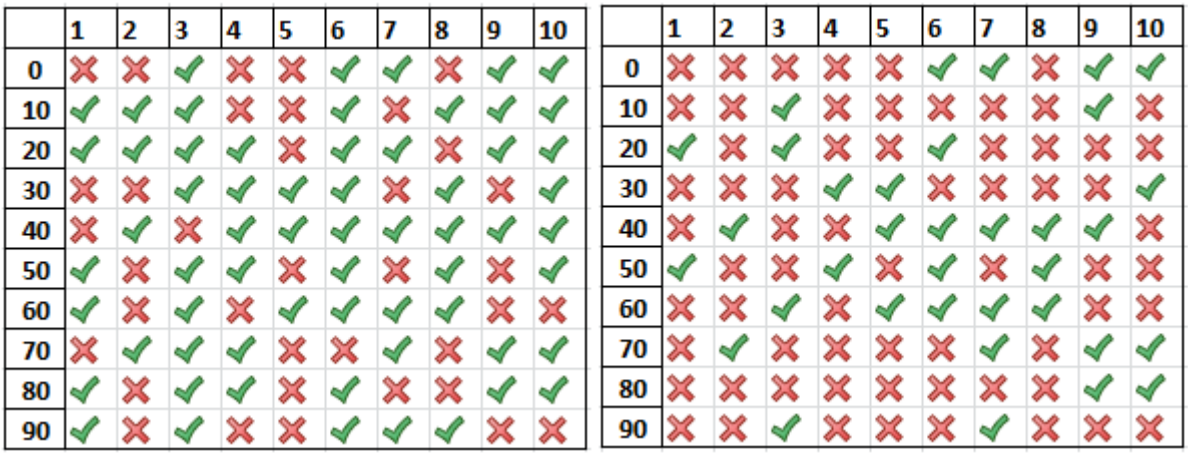


Figure 5.27: Change Chart Diagram for PARMS Results with and without Re-ranking

Initial processing of the question involves removing the stopwords from the question. Stemming is applied after term expansion using the WordNet ontology. A vector space model (Manning, Raghavan et al. 2008) has been used to calculate the distance between the question terms and terms of each document (verse). The English Translation of the Quran authored by Shakir has been used for this experiment.

```
<Question id="04" enabled="true">
  <Text>What are prohibited foods?</Text>
  <Answers>
    <Answer rank="1">002:173</Answer>
    <Answer rank="2">005:003</Answer>
    <Answer rank="3">006:145</Answer>
  </Answers>
  <Topic>Islamic Ruling:Food</Topic>
```

Figure 5.28: PARMS Model Question QID#04

The PARMS methodology does not identify 002:173 and 005:003 as potential answers. But 006:145 has been considered as a potential answer. The verse vocabulary was expanded by pronoun resolution using the QurAna ontology alone. The verse 006:145 was ranked 95th in order. But this order level increased when semantic knowledge from QurSim was utilised to

re-rank the answers. This brought verse 006:145 up to 85th place in the ranking order. Hence semantic knowledge represented in QurAna and QurSim adds value to the PARMS methodology.

Summary

Detailed experiments (see section 5.2) and an Evaluation results analysis (see section 5.3) of the proposed PARMS Methodology conclude that the Parallel-Corpus mode along with Multi-Stream semantic knowledge produces better results than individual corpus and less knowledge.

Chapter 6 - Conclusion & Future Work

In this thesis we have investigated question answering systems that are required to extract information from large, thematic corpuses, and in particular religious texts. The author has also developed a PARMS evaluation tool, which is specifically tailored and adapted to cope with the specific challenges of this genre. A characteristic of such systems is that they require a range of traditional statistical IR techniques including control-vocabulary generation, term-weight calculation and term-similarity measurement along with syntactic NLP methods such as tokenizing, stemming, parts-of-speech tagging and term expansion. These systems incorporate semantics in the shape of an ontology, or a formalised representation of technical terms and relations between these. We have applied our ideas to the Qur'an, utilising a combination of existing and novel resources and components. The results of the case study show that the best combination of specialised components in PARMS outperforms a standard IR keyword-based search.

Question-Answering on religious text is different from traditional Question-Answering on text collected from television programmes, news wires, conversations, novels and other like sources. The evaluation results of this research also support the hypothesis that additional semantic knowledge adds value to Question Answering Systems by helping in identifying correct answers. Traditional QA Systems reviewed in section 2.2.3 differ from QA systems aimed at Question-Answering involving the religious domain in the following aspects:

- **Nature of anticipated Questions:** Most QA systems aim at answering fact-based questions. Hence questions are aimed at some factual information. These questions are usually of the interrogatory type, i.e. 'What', 'Who', 'Why', 'How'. On the other hand, it is observed that questions in the religious domain are usually knowledge seeking as compared to probing facts.

- **Nature of Religious Corpus:** Religious texts usually exhibit peculiar characteristics that make them challenging but interesting. The Quranic text for example displays dense use of figurative language. The Quran contains rulings relating to religious concepts and viewpoints. Hence the user questions aim to probe certain aspects of a particular concept.

- **Nature of Expected Answer:** Alongside the nature of the question and corpus, the nature of the answer for a religious text is also likely to be different. An answer for a usual QA system is likely to be an extracted fact. But an answer to a religious question is usually a sub-section (paragraph) of the religious text itself, a Quranic verse for instance. Moreover the user expectation and consumption of this answer is also different. A 'factual' answer is usually consumed by the user *as it is*, whereas the answer to a religious question is firstly *analysed* by the user for its relatedness to their question. Later they *deduce knowledge* from the answer and determine its impact in the context of situation under consideration.

Recently the Quran – the core text for the Islam religion, has been considered as a potential subject for computing research in corpus linguistics (Dukes 2011), Text Analysis (Abdul-Baquee 2013) and natural language processing (Eric Atwell 2013). Research in these subject areas is cross-disciplinary as it involves knowledge of several domains and applications. Atwell states that "in general Natural Language Processing research involves Machine Learning from a domain-specific corpus of text documents". The following are a few attributes that make the source text corpus ideal for computational analysis:

- A text corpus should have no privacy constraints and costs involved. It should be available free for research purposes.
- A large community of experts for that corpus exists, which has developed 'linguistic tagging schemes' or ontologies for the domain.
- A large group of users exist, that can assist in evaluating the system and its results.
- A huge group of potential users exist, so the research in the domain has an impact

The Arabic Quran meets the above listed criteria in its fullness as there are around 1,300 years of documented work on the linguistics of the Arabic text of the Quran. Besides most features and resources mentioned above are valid for English translations of Quran also. In today's global population Muslim users of the Quran comprise "the largest user-group ever for a single text corpus" (Eric Atwell 2013). Out of this huge population, the audience user group for research on English translations of the Quran is still a huge group - the hundreds of millions of Muslims worldwide who use the English language. Quranic Arabic corpus linguistic research has produced ontological resources and tools such as the named entity ontology (Dukes 2011), semantic-concept "tags" (Abbas 2009), QurAna (Sharaf and Atwell 2012) and QurSim (Sharaf and Atwell 2012). Although these resources were developed for the Arabic Quran, they remain equally useful for research on English translations of the Quran as the verse references are the same regardless of the language of the Quran.

Due to the above stated differences the author was of the opinion that a traditional QA System requires adaptation for the religious domain. Besides, Monz states that the anticipated user of the Question-Answering System, the type of questions, the type of expected answers and the format in which available information is stored together determine the design of the Question Answering System (Monz 2003). However, the author believes that before this thesis, no fully-fledged research had been conducted with the aim of developing a methodology to achieve Question-Answering on text with the peculiar characteristics attributed to the religious domain. This thesis has developed the *PARMS* (**Par**allel Corpus **M**ulti **S**tream) methodology that attempts to fill this gap.

The PARMS Methodology is a novel method applying existing advanced IR (Information Retrieval) techniques combining them with NLP (Natural Language Processing) methods and additional semantic knowledge to implement QA (Question Answering) for a parallel corpus of English translations of the Quran – as a first case study. A Parallel Corpus involves the use of multiple forms of the same corpus where each form differs from others in a certain aspect, e.g. translations of a scripture from one language to another by different translators. It is vital to indicate that the PAMRS methodology uses multiple English translations of the Quran in multiple ways (parallel corpus or in isolation). Therefore this research has been carried out using one-step-away English translations of the original Arabic Quran. The rationale behind this is to make Quranic knowledge accessible to non-native speakers of Arabic, in the majority of Non-Arabic Muslim countries, e.g. Pakistan and Malaysia. Translations of the Quran inherit the complexity and peculiar characteristics of the Quranic language, which poses unique challenges to IR and NLP. Any QA System for the Quranic text that does not take the peculiar characteristics of English language translations into account is certain to be inadequate.

Generally, a QA system involves analysing texts extracted from a corpus to find a factoid or an entity that provides the answer. But it is also noticeable that the Islamic heritage emphasizes that answers to Islamic questions should be found in the original Arabic scripture. Hence, rather than generating factoid answers, PARMS analyses a question to pinpoint the source verses relevant to the question, which may be mapped onto the original Arabic script of the Quran or a translation of the Quran in any other language. Hence a significant difference in this work is the nature of the answers, i.e. they are not Answers in the classic sense.

Traditionally logic-based QA systems take a single text and aim to convert this to logical form such as predicate logic. Then inferencing is performed to deduce an answer. But this involves a single source, and alternative translations of a single *verse* would presumably map to the same logical representation. The PARMS methodology uses parallel English translations of the Quran, so it can calculate match against each translation, and/or against the set of translations. By contrast most QA systems work on a single source, so these cannot match against alternative translations. Hence PARMS matching to sub-section capitalises on the parallel corpus, whereas QA via logical inference would not.

Additional semantic knowledge can be referred as a *stream* of information related to a corpus. PARMS uses Multiple Streams of semantic knowledge including a general ontology (*WordNet*) and domain-specific ontologies (*QurTerms, QurAna, QurSim*). This additional knowledge has been used in embedded form for *Query Expansion*, *Corpus Enrichment* and *Answer Ranking*. The evaluation results of this research support the hypothesis that

additional semantic knowledge adds value to the Question Answering System by helping in identifying correct answers. But this assistance depends on the *nature* of the semantic knowledge and its application within the adopted methodology. By 'nature' I mean whether the semantic knowledge resource under consideration provides information about the corpus domain or knowledge about the corpus itself. For instance the QurTerms ontology used by the PARMS method helped decision-making by providing domain-related information, whereas knowledge derived from the QurAna ontology enriched the corpus itself. To remind the reader, the following are the novel contributions of this work:

I Question Answering System tailored to Religious Text
II PARMS Methodology
III Resource specific to QA of Religious Text – QurTerms
IV Model Question Answer Set – PARMS Model QA Set
V PARMS Evaluation tool

Potential Application areas of this research

This section introduces the potential applications of research contributions presented in this thesis. This section has been divided into sub-sections describing different aspects of these applications.

Methodology

The methodology of PARMS can be applied to NLP research on other texts that have peculiar characteristics as demonstrated by the Quranic text in this research. This could include parallel corpuses of other religious texts that have been translated into several international languages. This could include the Bible, Geeta, Grunth and similar religious texts. In fact, the Bible has a very similar structure to the Quran. The Bible has been translated in hundreds of languages and by several authors. Unlike the Quran, the Bible has different versions given that the specific Bible has a more or less authoritative status. This presents an opportunity for PARMS methodology to be applied on a parallel corpus of different versions of the Bible. As research in these areas is limited, there is a challenge in finding equivalent ontological resources for these research areas.

As discussed above, the PARMS QA system proposes a general methodology adaptable to similar research problems but it has direct application to other Islamic texts. This methodology can be applied to the freely available Islamic literature, involving hundreds of years of research on Hadith, Sunnah, and Exegesis. We believe that the approach adopted by the PARMS Methodology, i.e. combining IR & NLP baseline methods, enriched using

domain-specific semantic knowledge, could potentially have application in other relevant research fields that share attributes in common with the PARMS application domain. For example short texts such as Short Messaging Service and Micro-text such as Twitter Tweets could be an interesting start point.

Corpus

The Parallel Corpus used by the PARMS method comprises ten English translations of the Quran presented using XML technology. The PARMS Evaluation System enables use of this corpus in parallel or single-stream fashion. The technique of presenting complex data in XML format is an industry standard. It makes data more human readable and machine readable at the same time. XML representation of multiple Quranic translations in this research offers several benefits for similar research that aims at automated analysis of the Quran. This corpus will allow the following benefits for research in the area. It:

- Reduces complexity of data representation
- Backs reusability
- Enables portability
- Allows database independence
- Encourages self-sufficient system development
- Enables quicker and easier data parsing
- Increases speed performance impact

Moreover, the diversity in author selection provides variation in writing style, English language period, cultural background, and scholarly level of authors. We propose that this corpus can be used for wider research in:

- Translation Studies: to understand the complexity of language of religious texts
- Linguistic Studies: to study how English has changed over time

QurTerms

The PARMS QA system uses the QurTerms ontology, which in effect can be considered an Islamic WordNet plus gloss resource. QurTerms holds useful semantic knowledge of three basic types.

- Firstly, QurTerms provides different *linguistic representations* of the same Islamic terms as used by different authors, translators, scholars and the general public.
- Secondly, QurTerms provides *meanings* for those Islamic terms which do not belong to contemporary English. Hence their meanings are not available in glosses and dictionaries.

- Lastly, QurTerms provides a *thesaurus* for those Islamic terms that are commonly used and take different forms depending on their context of use. No thesaurus of these words is usually available in the WordNet ontology.

The QurTerms ontology uses XML representation. Hence it carries all the advantages in-built with this representation, as discussed above. The PARMS QA System uses this resource to resolve such Islamic Terms which occur in the natural language of speakers referring to Islamic concepts either for asking questions, or communicating knowledge. The proposed technique can be applied to further works that involve Islamic terms for which the semantic knowledge does not exist elsewhere. Besides, this technique can be adapted to develop such resources for other religious texts. For example for the Bible this resource can be used for Biblical terms instead of Quranic terms.

Application

The PARMS QA System as an application is useful for Islamic scholars who need to make quick reference to the Quran for research and investigation before giving out advice or reference statements. Similarly these resources can be made available to students of religion who consult the Quran from the perspective of their own study. Finally these can be embedded in applications that aim at providing decision support mechanisms to the general public.

Question Set

The PARMS QA System has been evaluated using 100 Model Questions. These questions have been collected and extracted from several sources after extensive research through similar available resources. The selection criterion for each question was aimed at testing a particular aspect of the system. But the extraction method and rationale behind the selections can be adapted to build on this set for future research. Hence the we propose that this set of Questions can be considered as a *Gold standard* for evaluating a Question Answering system's efficiency with the Quranic text.

Although these questions follow XML representation schema, they vary widely in their type and nature. The rational and assumptions behind the decisions are presented in detail in Chapter 6. The question set will be made available for reuse under a General Public Licence. Hence the *Gold Standard Question Set* will be freely available and can be used to continue research in the area.

Quranic NLP Resources

The resources, techniques and methods proposed in this thesis can be considered a starting point for NLP resources specific to the Quran. The Quran is a vital book as it is a core text

which contributes to the lives of millions of people today. There is a need for more research on this book so as to make the latent knowledge available and accessible. The Quranic NLP resources produced (QurTerms) and existing Quranic NLP resources (QurAna, QurSim) used can be considered initial step in this direction.

PARMS QA System

To the best of our knowledge, the PARMS QA system is the first ever attempt at Question Answering on the Quran using the novel domain-specific ontologies, QurAna & QurSim. We suggest that the PARMS QA System can also be considered a *Benchmark System* for Question Answering on the Quran. This software and connected computational resources will be made available under a General Public Licence. Hence it will be freely available and will be able to be used to compare further research in the area.

Besides the above application of the work presented in this thesis, there is potential to enhance the PARMS Methodology further in future. The following could be indicators of further research in this direction:

- A comparison between the semantics of both the question and the potential answers can be performed to improve the results. This process can be applied within different phases of the methodology, e.g. answer extraction or answer re-ranking to improve the overall results.

- Commentaries on the Quran such as Tafsir Ibn-Kathir are a useful source of additional knowledge about the Quran. These sources could be incorporated within the method at different levels, e.g. parallel-corpus design, query expansion, answer extraction, corpus enrichment or answer re-ranking.

- There is on-going work on tagging the Arabic Quran from a linguistic perspective. The output of these works is likely to produce additional knowledge about the Quran. There is potential to incorporate these resources and similar works within the PARMS methodology once they are made available, to enhance results.

- There are certain topics of interest in the Quran that attract frequent user questions. These topics are very complex and involve several inter-locking concepts, for instance, rulings on the concept of Mahram (un-marriageable kin) and Inheritance. There is a potential to apply improved knowledge representation techniques to Quranic verses that provide information on these concepts. A dedicated method could be developed for knowledge inference from these representations to find answers to user questions related to these topics.

- The work in this thesis covers the Quran as a whole, which makes it a huge domain for further work as suggested above. This scope could be restricted to a limited sub-section of the Quran. For instance, detailed experimentation could be performed with the focus on certain chapters, or certain topics of interest within the Quran and later enhanced to encapsulate the Quran as a whole.

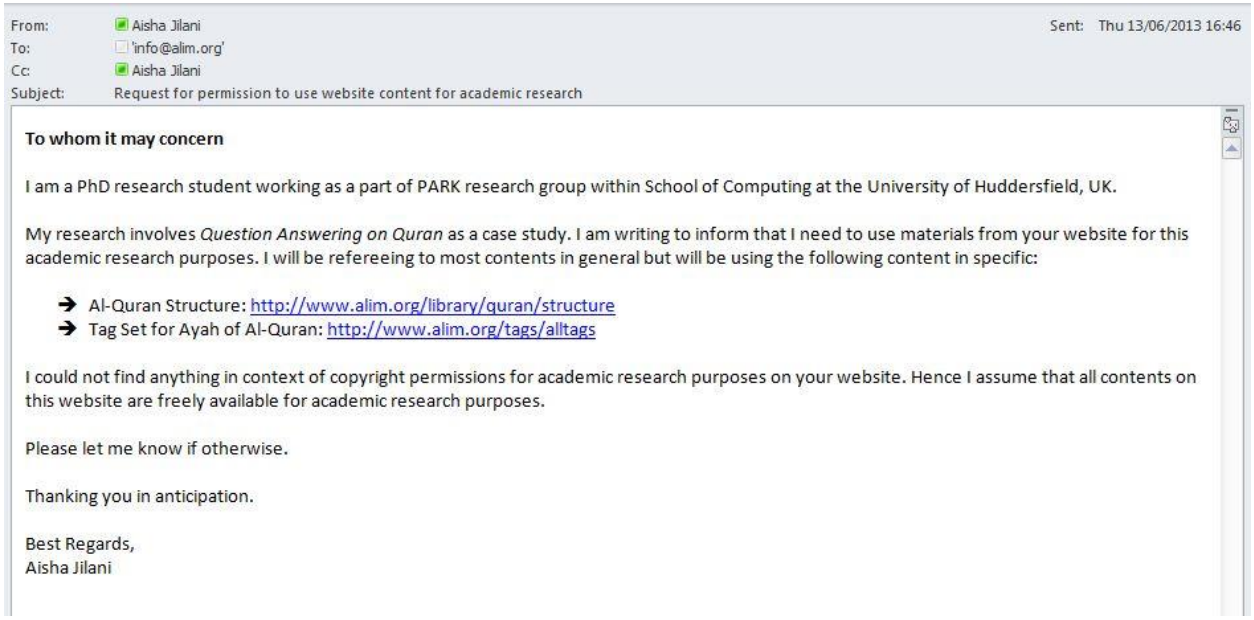
In essence, an ideal Question Answering System for a religious text has to deal with several problems pertinent to the nature and structure of the religious language alongside traditional research issues around Information Retrieval (IR) and Natural Language Processing (NLP) domain. Challenges posed by the poetic religious language of the Quran and the additional constraint that the Classical Arabic source text must not be tampered with require a method that may provide answers to English language questions that are true to the source too. The research presented in this thesis has attempted to meet these challenges.

Appendices

Appendix 1 – Copyright Permissions

Alim.org

Following is the email sent through to website owners to seek permission for using materials from the website.



From: Aisha Jilani
To: 'info@alim.org'
Cc: Aisha Jilani
Subject: Request for permission to use website content for academic research
Sent: Thu 13/06/2013 16:46

To whom it may concern

I am a PhD research student working as a part of PARK research group within School of Computing at the University of Huddersfield, UK.

My research involves *Question Answering on Quran* as a case study. I am writing to inform that I need to use materials from your website for this academic research purposes. I will be refereeing to most contents in general but will be using the following content in specific:

- Al-Quran Structure: http://www.alim.org/library/quran/structure
- Tag Set for Ayah of Al-Quran: http://www.alim.org/tags/alltags

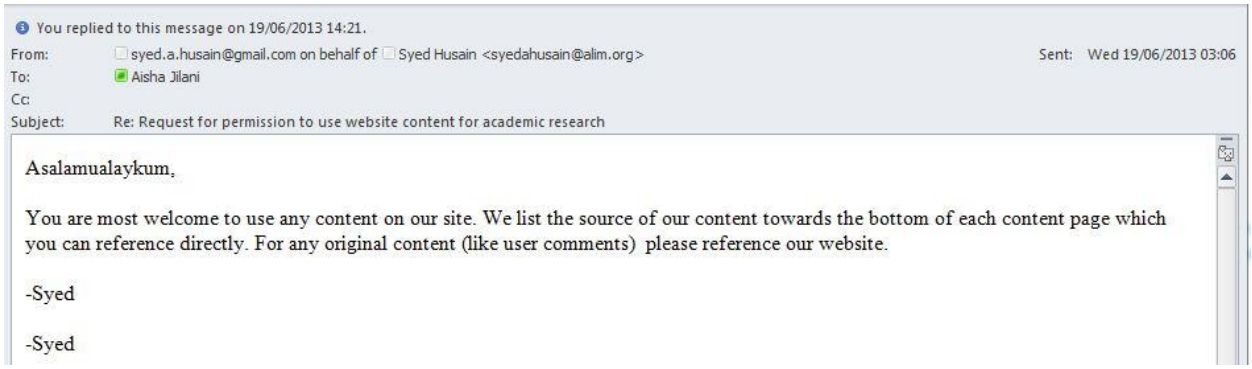
I could not find anything in context of copyright permissions for academic research purposes on your website. Hence I assume that all contents on this website are freely available for academic research purposes.

Please let me know if otherwise.

Thanking you in anticipation.

Best Regards,
Aisha Jilani

Following is the email reply to the email through to website owners to seek permission for using materials from the website.



You replied to this message on 19/06/2013 14:21.
From: syed.a.husain@gmail.com on behalf of Syed Husain <syedahusain@alim.org>
To: Aisha Jilani
Cc:
Subject: Re: Request for permission to use website content for academic research
Sent: Wed 19/06/2013 03:06

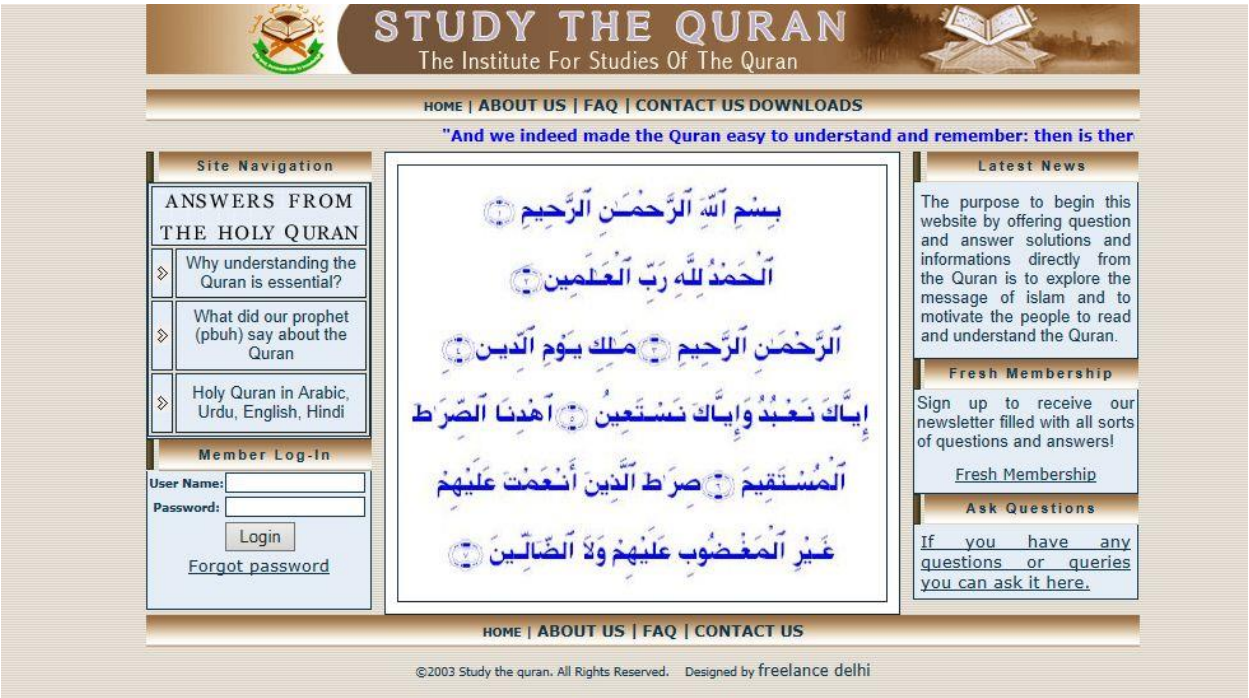
Asalamualaykum,

You are most welcome to use any content on our site. We list the source of our content towards the bottom of each content page which you can reference directly. For any original content (like user comments) please reference our website.

-Syed

-Syed

StudytheQuran.com



Following is the email sent through to website owners to seek permission for using materials from the website.

From: Aisha Jilani
Sent: Thu 21/02/2013 11:33
To: 'info@studythequran.com'
Cc: Aisha Jilani; Aisha Jilani U0416820; 'aisha.jilani@gmail.com'
Subject: Request for permission to use website content for academic research

To whom it may concern

I am a PhD research student working as a part of PARK research group within School of Computing at the University of Huddersfield UK.

My research involves *Question Answering on Quran* as a case study. I am writing to inform that I need to use materials from your website for this academic research purposes. I will be refereeing to most contents in general but will be using the following content in specific:

- Extracts from the Question Answer pairs database available at: http://studythequran.com/answers_from_the_holyquran.php
- Subject Category List available at: http://studythequran.com/display.php?sub_id=89

It does not say anything in context of copyrights on your website so I assume that all contents on this website are freely available for academic research purposes.

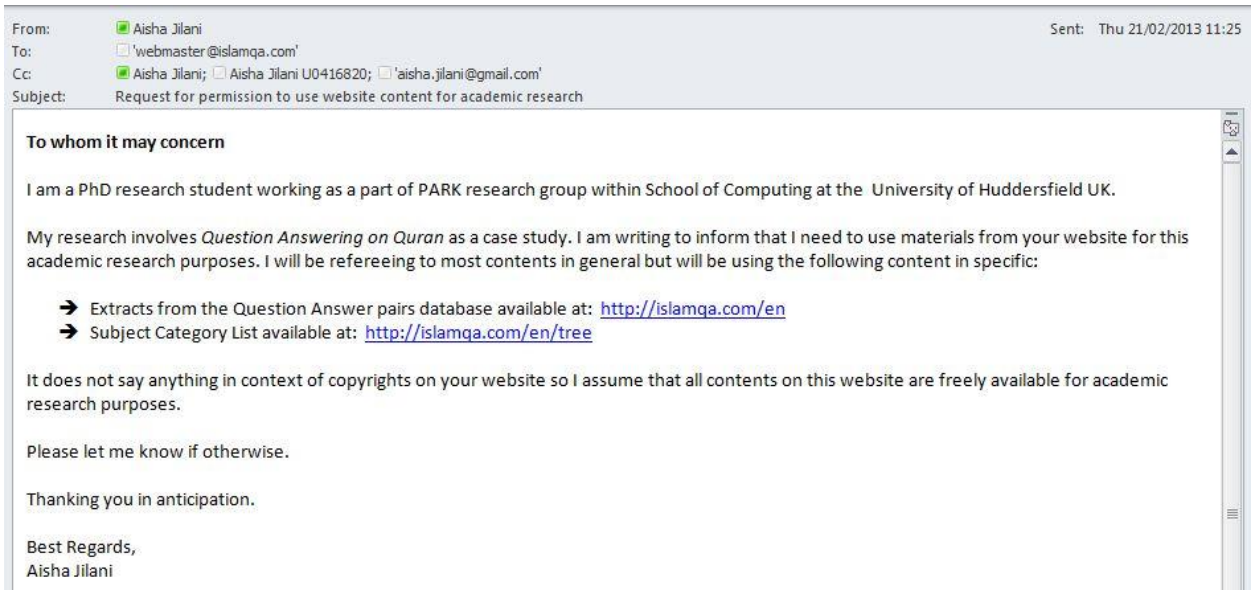
Please let me know if otherwise.

Thanking you in anticipation.

Best Regards,
Aisha Jilani

IslamQA.info

Following is the email sent through to website owners to seek permission for using materials from the website.



From: Aisha Jilani
To: 'webmaster@islamqa.com'
Cc: Aisha Jilani; Aisha Jilani U0416820; 'aisha.jilani@gmail.com'
Subject: Request for permission to use website content for academic research
Sent: Thu 21/02/2013 11:25

To whom it may concern

I am a PhD research student working as a part of PARK research group within School of Computing at the University of Huddersfield UK.

My research involves *Question Answering on Quran* as a case study. I am writing to inform that I need to use materials from your website for this academic research purposes. I will be refereeing to most contents in general but will be using the following content in specific:

- Extracts from the Question Answer pairs database available at: http://islamqa.com/en
- Subject Category List available at: http://islamqa.com/en/tree

It does not say anything in context of copyrights on your website so I assume that all contents on this website are freely available for academic research purposes.

Please let me know if otherwise.

Thanking you in anticipation.

Best Regards,
Aisha Jilani

Following is the email reply to the email through to website owners to seek permission for using materials from the website.

From: Website.IslamQA.com [mailto:fatwa@islamqa.com] وجواب سؤال الإسلام موقع
Sent: 23 February 2013 14:41
To: Aisha Jilani
Subject: Answer to your question

Assalamu alaikum,

We ask Allaah to make you steadfast in your efforts to fulfil the duty of calling people to Allaah and in your keenness to spread good. Your love of good and your efforts to spread it indicate that you are a good person, in sha Allah. We ask Allaah that you will be included in His words (interpretation of the meaning):

"You (true believers in Islamic Monotheism, and real followers of Prophet Muhammad and his Sunnah) are the best of peoples ever raised up for mankind; you enjoin Al-Ma'roof (i.e. Islamic Monotheism and all that Islam has ordained) and forbid Al-Munkar (polytheism, disbelief and all that Islam has forbidden), and you believe in Allaah"

Permission is granted to all to take material from this site, subject to the following two conditions: (1) Material used must be attributed to www.islam-qa.com , (2) Material must be reproduced faithfully and without alteration or omission. And Allaah is the Source of strength.

Islamqa.com

--

Appendix 2 – XML Schemas

Question Set

```
<?xml version="1.0" encoding="UTF-8"?>

<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
    <xsd:element name="QuestionSet">
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element name="Questions">
                    <xsd:complexType>
                        <xsd:sequence>
                            <xsd:element name="Question"
                                         type="QAQuestion"
                                         minOccurs ="1"
                                         maxOccurs="unbounded" />
                        </xsd:sequence>
                    </xsd:complexType>
                </xsd:element>
            </xsd:sequence>
        </xsd:complexType>
    </xsd:element>
    <xsd:complexType name="QAQuestion">
        <xsd:sequence>
            <xsd:element name="Text"
                         type="xsd:string" />
            <xsd:element name="Answers">
                <xsd:complexType>
                    <xsd:sequence>
                        <xsd:element name="Answer"
                                     type="QAAnswer"
                                     minOccurs ="1"
                                     maxOccurs="unbounded" />
                    </xsd:sequence>
                </xsd:complexType>
            </xsd:element>
            <xsd:element name="Topic" type="xsd:string" />
            <xsd:element name="SourceContext" type="QASourceContext" />
        </xsd:sequence>
        <xsd:attribute name="id" type="xsd:string" use="required"/>
        <xsd:attribute name="enabled" type="xsd:boolean" use="required"/>
    </xsd:complexType>
    <xsd:complexType name="QAAnswer">
        <xsd:simpleContent>
          <xsd:extension base="xsd:string">
            <xsd:attribute name="rank" type="xsd:int" use="required" />
          </xsd:extension>
        </xsd:simpleContent>
    </xsd:complexType>
    <xsd:complexType name="QASourceContext">
        <xsd:sequence>
           <xsd:element name="Source" type="xsd:string" />
           <xsd:element name="DateTaken" type="xsd:date" />
        </xsd:sequence>
    </xsd:complexType>

</xsd:schema>
```

QurTerms Ontology

```
<!-- xml type for Quranic terms -->
<xs:element name="QuranicTermsList">
  <xs:complexType>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" name="QuranicTerm">
        <xs:complexType>
          <xs:sequence>
            <xs:element name="term" type="xs:string" />
            <xs:element name="meanings">
              <xs:complexType>
                <xs:sequence>
                  <xs:element maxOccurs="unbounded" name="meaning" type="xs:string" />
                </xs:sequence>
              </xs:complexType>
            </xs:element>
          </xs:sequence>
          <xs:attribute name="id" type="xs:unsignedShort" use="required" />
        </xs:complexType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

Appendix 3 – PARMS Input

QurTerms

```
<?xml version="1.0" encoding="utf-8"?>
<QuranicTermsList>
  <QuranicTerm id="1">
    <term>Allaah | Alaah | Allah</term>
    <meanings>
      <meaning>Allah</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="2">
    <term>Quraan | Qur'aan | Koran | Qur'an | AlQuran | Mushaf | Mus-haf</term>
    <meanings>
      <meaning>quran</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="3">
    <term>Soorah | Surah | Sura | Surat | Soorat</term>
    <meanings>
      <meaning>sura</meaning>
         <meaning>surah</meaning>
         <meaning>chapter</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="4">
    <term>Ayah | Ayat | Ayaah | Verse</term>
    <meanings>
      <meaning>verse</meaning>
         <meaning>message</meaning>
         <meaning>revelation</meaning>
         <meaning>sign</meaning>
         <meaning>communication</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="5">
    <term>Salat | Solat | As-Salat | Salah | As-Salaah | Namaz</term>
    <meanings>
      <meaning>prayer</meaning>
         <meaning>As-Salat</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="6">
    <term>Wudu | Wudoo | Wazu</term>
    <meanings>
      <meaning>Ablution</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="7">
    <term>Tayammum</term>
    <meanings>
      <meaning>dry ablution</meaning>
         <meaning>wipe clean with pure earth dust sand</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="8">
    <term>Iblis | Iblees | Satan | Shaytan | Shaytaan | Shaitan</term>
    <meanings>
      <meaning>devil</meaning>
```

```
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="9">
    <term>Talaq | Tallaq | Taalaq | Talaaq</term>
    <meanings>
      <meaning>divorce</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="10">
    <term>Hajj</term>
    <meanings>
      <meaning>pilgrimage</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="11">
    <term>Umrah | Umra</term>
    <meanings>
      <meaning>pay a visit</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="12">
    <term>Fard</term>
    <meanings>
      <meaning>duty</meaning>
      <meaning>incumbent</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="13">
    <term>du'aa' | Duaa</term>
    <meanings>
      <meaning>supplication</meaning>
         <meaning>prayers</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="14">
    <term>Kafir | Kaafar</term>
    <meanings>
      <meaning>disbelievers</meaning>
         <meaning>unbelievers</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="15">
    <term>Hijaab | Hijab | burka | burqa</term>
    <meanings>
      <meaning>veils</meaning>
         <meaning>cloaks</meaning>
         <meaning>let down upon over-garments</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="16">
    <term>Ramadan | Ramadhan | Ramazan | Ramzan | Saum | As-Saum | Sayyam | Roza </term>
    <meanings>
      <meaning>Ramadan</meaning>
         <meaning>Fasting</meaning>
         <meaning>To Fast</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="17">
    <term>Zakat | Zakaat | zakaah</term>
    <meanings>
         <meaning>charity</meaning>
         <meaning>poor-rate</meaning>
    </meanings>
  </QuranicTerm>
```

```
<QuranicTerm id="18">
  <term>Ribba | Riba | Ribah</term>
  <meanings>
    <meaning>usury</meaning>
       <meaning>interest</meaning>
  </meanings>
</QuranicTerm>
<QuranicTerm id="19">
  <term>Halal | Hallal | Al-Halal</term>
  <meanings>
    <meaning>lawful</meaning>
       <meaning>allowed</meaning>
       <meaning>permitted</meaning>
  </meanings>
</QuranicTerm>
<QuranicTerm id="20">
  <term>Haram | Harram | Haraam</term>
  <meanings>
    <meaning>forbidden</meaning>
       <meaning>prohibited</meaning>
       <meaning>unlawful</meaning>
  </meanings>
</QuranicTerm>
<QuranicTerm id="21">
  <term>Idaah | Idat | Idaat | iddat | Iddah</term>
  <meanings>
    <meaning>prescribed period or days</meaning>
       <meaning>period or term or time to reckon or wait</meaning>
  </meanings>
</QuranicTerm>
     <QuranicTerm id="22">
  <term>Qisas | Al-Qisas | Qusas | Diya | Fiddiya | Qasaas | Fidyah</term>
  <meanings>
    <meaning>Blood money</meaning>
       <meaning>Blood ransom</meaning>
       <meaning>Blood compensation</meaning>
       <meaning>Innocent Killing payment</meaning>
       <meaning>Retaliation</meaning>
       <meaning>Equitable retaliation</meaning>
       <meaning>Retaliation of blood</meaning>
       <meaning>Law of equality in punishment</meaning>
  </meanings>
</QuranicTerm>
<QuranicTerm id="23">
  <term>Mahram | Mehrum | Mahraam</term>
  <meanings>
    <meaning>Forbidden or Prohibited for Marriage</meaning>
       <meaning>Women that cannot ne taken in Marriage</meaning>
  </meanings>
</QuranicTerm>
<QuranicTerm id="24">
  <term>Mahr | Meher | Maher</term>
  <meanings>
    <meaning>obligatory bridal money</meaning>
       <meaning>dowries</meaning>
       <meaning>dowry</meaning>
       <meaning>dower</meaning>
  </meanings>
</QuranicTerm>
<QuranicTerm id="25">
  <term>Khola | Khola' | Khula | Al-Khul</term>
  <meanings>
       <meaning>set free from bond of marriage</meaning>
```

```
      <meaning>Inability to repay dower and the custody of children</meaning> <!-- Taken
from IslamCity at http://www.islamicity.com/qa/action.lasso.asp?-db=services&-lay=Ask&-
op=eq&number=4546&-format=detailpop.shtml&-find -->
         <meaning>wife willingly gives up to become free</meaning>
         <meaning>ransom to set free</meaning>
         <meaning>give up to get freedom</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="26">
    <term>Makrooh | Makruh | Makroh | Makrouh</term>
    <meanings>
         <meaning>Hateful</meaning>
      <meaning>Disliked</meaning>
         <meaning>detestable</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="27">
    <term>Sadqa | Sadaqah | Sadka</term>
    <meanings>
         <meaning>Sadaqah</meaning>
      <meaning>voluntary charity</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="28">
    <term>Jannat | Jinnah | Junat | Junnah</term>
    <meanings>
         <meaning>Heaven</meaning>
      <meaning>Paradise</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="29">
    <term>Zabiha | Zibah | Zabeeha </term>
    <meanings>
         <meaning>to butcher animal</meaning>
         <meaning>to slaughter in Islamic way</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="30">
    <term>Tobah | Taoba | Toba | Tobba | Tobaa</term>
    <meanings>
      <meaning>to repent</meaning>
      <meaning>repentance</meaning>
         <meaning>to beg forgiveness</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="31">
    <term>Niyyah | Nieh | Niyat </term>
    <meanings>
      <meaning>intention</meaning>
      <meaning>intention for fasting</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="32">
    <term>Zihar | Thihaar | Dhihar </term>
    <meanings>
      <meaning>Divorce by calling wife your mother</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="33">
    <term>Kaffarah | Kafarah | Kafara </term>
    <meanings>
      <meaning>Expiation</meaning>
         <meaning>Compensate</meaning>
    </meanings>
```

```
 </QuranicTerm>
<QuranicTerm id="34">
   <term>Ahram | Ihram | Ehram </term>
   <meanings>
     <meaning>Hajj Dress</meaning>
        <meaning>Dressing to perform Hajj</meaning>
   </meanings>
 </QuranicTerm>
<QuranicTerm id="35">
   <term>Al-Muttaqun | Muttaqeen | Muttaqi </term>
   <meanings>
     <meaning>Pious</meaning>
        <meaning>Righteous</meaning>
        <meaning>those who stay from evil</meaning>
        <meaning>follower of the right path</meaning>
   </meanings>
 </QuranicTerm>
 <QuranicTerm id="36">
   <term>Orah | Awrah | Owra </term>
   <meanings>
     <meaning>Private body parts</meaning>
        <meaning>Genitals</meaning>
        <meaning>Private body parts which are required to be covered in public</meaning>
   </meanings>
 </QuranicTerm>
<QuranicTerm id="37">
   <term>Salam | Aslam-o-Alikum | Salamualaikum | Salaam | As-Salamu 'Alaikum </term>
   <meanings>
     <meaning>Muslim Greetings</meaning>
        <meaning>to greet a muslim</meaning>
        <meaning>Islamic way of Saluting</meaning>
        <meaning>Salutation</meaning>
   </meanings>
 </QuranicTerm>
 <QuranicTerm id="38">
   <term>Ramzan | Ramadan | Ramadhan | Ramadaan </term>
   <meanings>
     <meaning>Month of Islamic Lunar Calender</meaning>
        <meaning>Month of Fasting</meaning>
   </meanings>
 </QuranicTerm>
 <QuranicTerm id="39">
   <term>Azan | Aadan | Adhan | Aadaan </term>
   <meanings>
     <meaning>Call for Prayer</meaning>
        <meaning>Month of Fasting</meaning>
   </meanings>
 </QuranicTerm>
 <QuranicTerm id="40">
   <term>Fatwa | Fatva | Futwa </term>
   <meanings>
     <meaning>Religious Ruling</meaning>
        <meaning>Scholar Answer</meaning>
        <meaning>Religious decision</meaning>
        <meaning>Scholar Advice</meaning>
   </meanings>
 </QuranicTerm>
 <QuranicTerm id="41">
   <term> Sarh | As-Sarh </term>
   <meanings>
     <meaning>a glass surface with water underneath it</meaning>
        <meaning>Crystal surface with water under it</meaning>
        <meaning>Lake of Water</meaning>
        <meaning>Expanse of Water</meaning>
```

```
        <meaning>Pool of Water</meaning>
        <meaning>Floor of Water</meaning>
    </meanings>
  </QuranicTerm>
  <QuranicTerm id="42">
    <term> Alamin | Alameen | Raubulalmeen | RabeAlamin </term>
    <meanings>
      <meaning>Lord of mankind, jinns and all that exists</meaning>
        <meaning>Lord of the Universe</meaning>
        <meaning>Lord of the Worlds</meaning>
    </meanings>
  </QuranicTerm>
</QuranicTermsList>
```

PARMS Model Q&A Set

```
<?xml version="1.0" encoding="utf-8"?>

 <!-- Number of FAQ's = 40 -->
 <!-- Number of Online User Queries = 48 -->
 <!-- Number of Offline User Queries = 08 -->
 <!-- Number of Derived User Queries = 04 -->

 <!-- Number of Questions on Topic: Quran = 02-->
 <!-- Number of Questions on Topic: History = 04-->
 <!-- Number of Questions on Topic: Social Norms = 07-->
 <!-- Number of Questions on Topic: Faith = 14-->
 <!-- Number of Questions on Topic: Rituals = 17-->
 <!-- Number of Questions on Topic: Code of Conduct = 25-->
 <!-- Number of Questions on Topic: Islamic Ruling = 31-->

<QuestionSet>
 <Questions>
    <Question id="01" enabled="true">
      <Text>Assalamualaikum..  Is  crab  halal  or  haram  so  as  shrimps  am
confused...Thanks</Text>
      <Answers>
        <Answer rank="1">005:096</Answer>
      </Answers>
      <Topic>Islamic Ruling:Food</Topic>
      <SourceContext>
        <Source><![CDATA[    http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=34018&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> 2010-02-15 </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="02" enabled="true">
      <Text> I am the head of a household, with a wife and two children. My salary barely
covers the necessities of life, and I do not have any other income. I worked in one of
the Gulf countries for 4 years, and I saved an amount of money which I deposited in one
of the Islamic banks to provide me with an income to help me deal with different burdens
of life. My salary and this income from this investment are just sufficient for me and my
family. Am I obliged to spend some of this deposited money on Hajj expenses, and am I
obliged to go for Hajj in the light of these circumstances?
 Please note that if I spend this amount from my bank account on Hajj expenses, this will
affect my monthly income and will make things very difficult for me financially. What do
you advise me to do?
</Text>
      <Answers>
        <Answer rank="1">003:097</Answer>
      </Answers>
      <Topic>Rituals:Pilgrimage</Topic>
      <SourceContext>
        <Source> http://islamqa.info/en/ref/36437</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="03" enabled="true">
      <Text>A person died, leaving behind a mother, father and four brothers. How should
the inheritance be divided?</Text>
      <Answers>
        <Answer rank="1">004:011</Answer>
```

```
    </Answers>
    <Topic>Code of Conduct:Inheritance</Topic>
    <SourceContext>
      <Source>http://islamqa.info/en/ref/13632/inheritance</Source>
      <DateTaken>2013-03-22</DateTaken>
      <QuestionDate> Unknown </QuestionDate>
      <QuestionType> Online User Query </QuestionType>
    </SourceContext>
  </Question>
  <Question id="04" enabled="true">
    <Text>What are prohibited foods?</Text>
    <Answers>
      <Answer rank="1">002:173</Answer>
      <Answer rank="2">005:003</Answer>
      <Answer rank="3">006:145</Answer>
    </Answers>
    <Topic>Islamic Ruling:Food</Topic>
    <SourceContext>
      <Source>http://www.studythequran.com/ShowQandA.php?qa_id=70</Source>
      <DateTaken>2012-11-22</DateTaken>
      <QuestionDate> Unknown </QuestionDate>
      <QuestionType> FAQ </QuestionType>
    </SourceContext>
  </Question>
  <Question id="05" enabled="true">
    <Text>What is the criterion for lawful food?</Text>
    <Answers>
      <Answer rank="1">002:172</Answer>
      <Answer rank="2">005:088</Answer>
      <Answer rank="3">016:114</Answer>
      <Answer rank="4">023:051</Answer>
    </Answers>
    <Topic>Islamic Ruling:Food</Topic>
    <SourceContext>
      <Source>http://www.studythequran.com/ShowQandA.php?qa_id=69</Source>
      <DateTaken>2012-11-22</DateTaken>
      <QuestionDate> Unknown </QuestionDate>
      <QuestionType> FAQ </QuestionType>
    </SourceContext>
  </Question>
  <Question id="06" enabled="true">
    <Text>What is the difference between usury and trade?</Text>
    <Answers>
      <Answer rank="1">002:275</Answer>
    </Answers>
    <Topic>Islamic Ruling:Usury</Topic>
    <SourceContext>
      <Source>http://www.studythequran.com/ShowQandA.php?qa_id=285</Source>
      <DateTaken>2012-11-22</DateTaken>
      <QuestionDate> Unknown </QuestionDate>
      <QuestionType> FAQ </QuestionType>
    </SourceContext>
  </Question>
  <Question id="07" enabled="true">
    <Text>What are Allah's instructions to mankind for his/her parents?</Text>
    <Answers>
      <Answer rank="1">046:015</Answer>
      <Answer rank="2">031:014</Answer>
    </Answers>
    <Topic>Code of Conduct:Family</Topic>
    <SourceContext>
      <Source>http://www.studythequran.com/ShowQandA.php?qa_id=162</Source>
      <DateTaken>2012-11-22</DateTaken>
      <QuestionDate> Unknown </QuestionDate>
```

```
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="08" enabled="true">
      <Text>What special regards we are required to give to our parents in their old
age?</Text>
      <Answers>
        <Answer rank="1">017:023</Answer>
      </Answers>
      <Topic>Code of Conduct:Family</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=164</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="09" enabled="true">
      <Text>What is the prescribed waiting period for a widow?</Text>
      <Answers>
        <Answer rank="1">002:234</Answer>
      </Answers>
      <Topic>Code of Conduct:Waiting Period</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=126</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="10" enabled="true">
      <Text>Does the widow woman have right to lodge in her late husband's house?</Text>
      <Answers>
        <Answer rank="1">002:240</Answer>
      </Answers>
      <Topic>Code of Conduct:Family</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=262</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="11" enabled="true">
      <Text>What are the mandatory acts of ablution?</Text>
      <Answers>
        <Answer rank="1">005:006</Answer>
      </Answers>
      <Topic>Islamic Ruling:Ablution</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=1</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="12" enabled="true">
      <Text>Backbiting and discovering secret of others is too common to be noticed or
considered as sin. How severe is this sin in the sight of Allah?</Text>
      <Answers>
        <Answer rank="1">049:012</Answer>
        <Answer rank="2">104:001</Answer>
      </Answers>
      <Topic>Social Norms:Backbiting</Topic>
```

```
      <SourceContext>
        <Source><![CDATA[ Offline ]]></Source>
        <DateTaken>2012-10-26</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Offline User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="13" enabled="true">
      <Text>In the story of Sulaymaan and the Queen of Saba' can Sarh be classed as an
architectural wonder or was is it just a miricle? </Text>
      <Answers>
        <Answer rank="1">027:044</Answer>
      </Answers>
      <Topic>History:Prophets</Topic>
      <SourceContext>
        <Source><![CDATA[ Offline ]]></Source>
        <DateTaken>2012-10-26</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Offline User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="14" enabled="true">
      <Text>Is any order of preference prescribed for spending our wealth?</Text>
      <Answers>
        <Answer rank="1">002:215</Answer>
      </Answers>
      <Topic>Code of Conduct:Charity</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=10</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="15" enabled="true">
      <Text>How much should be given as alms?</Text>
      <Answers>
        <Answer rank="1">002:219</Answer>
      </Answers>
      <Topic>Code of Conduct:Charity</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=11</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate>Unknown</QuestionDate>
        <QuestionType> Derived User Query </QuestionType>

      </SourceContext>
    </Question>
    <Question id="16" enabled="true">
      <Text>Which one is closer to God: giving charity publicly or secretly?</Text>
      <Answers>
        <Answer rank="1">002:271</Answer>
      </Answers>
      <Topic>Code of Conduct:Charity</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=16</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="17" enabled="true">
      <Text>Is friendship with non-believer forbidden?</Text>
      <Answers>
```

```
        <Answer rank="1">004:144</Answer>
        <Answer rank="2">005:051</Answer>
        <Answer rank="2">060:013</Answer>
      </Answers>
      <Topic>Social Norms:Dealing with Non-Muslims</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=35</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="18" enabled="true">
      <Text>After I divorce my wife, is her mother regarded as a mahram?</Text>
      <Answers>
        <Answer rank="1">004:023</Answer>
      </Answers>
      <Topic>Islamic Ruling:Unmarriageable Kin</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/95757</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="19" enabled="true">
      <Text>Does Quran provide any evidence on the scientific fact that Sun and Moon
follow a certain law in their motion? </Text>
      <Answers>
        <Answer rank="1">013:002</Answer>
        <Answer rank="2">031:029</Answer>
        <Answer rank="3">035:013</Answer>
        <Answer rank="4">036:040</Answer>
        <Answer rank="5">039:005</Answer>
      </Answers>
      <Topic>Quran</Topic>
      <SourceContext>
        <Source><![CDATA[ Offline ]]></Source>
        <DateTaken>2012-10-26</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Offline User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="20" enabled="true">
      <Text>Husband want to marry his wife sister daughter (means wife niece). Can he
marry his wife niece and having at the same time both women in nikkah?</Text>
      <Answers>
        <Answer rank="1">004:023</Answer>
      </Answers>
      <Topic>Islamic Ruling:Unmarriageable Kin</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/147367</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="21" enabled="true">
      <Text>Does the wife have the right to financial compensation from the husband if
she breastfeeds his children?</Text>
      <Answers>
        <Answer rank="1">065:006</Answer>
        <Answer rank="2">002:233</Answer>
      </Answers>
```

```
      <Topic>Code of Conduct:Breastfeeding</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/130116</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="22" enabled="true">
      <Text>Is it permissible or not for a woman to uncover her breast in front of her
mahrams in order to breastfeed her child, if there is no fear of fitnah?</Text>
      <Answers>
        <Answer rank="1">024:031</Answer>
      </Answers>
      <Topic>Code of Conduct:Breastfeeding</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/113287</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="23" enabled="true">
      <Text>Can I pray after doing tayammum in the case of janabah on intensely cold days
knowing that I do not have the ability to purify myself immediately? In addition to that,
I get ill because of the cold that affects my back, which affects me a great deal</Text>
      <Answers>
        <Answer rank="1">005:006</Answer>
      </Answers>
      <Topic>Islamic Ruling:Ablution</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/70507</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="24" enabled="true">
      <Text>What is the alternate if water is not found for ablution?</Text>
      <Answers>
        <Answer rank="1">005:006</Answer>
      </Answers>
      <Topic>Islamic Ruling:Ablution</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=1</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Derived User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="25" enabled="true">
      <Text>What will be the end of those who continue to practise usury?</Text>
      <Answers>
        <Answer rank="1">004:161</Answer>
      </Answers>
      <Topic>Islamic Ruling:Usury</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=285</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="26" enabled="true">
```

```
      <Text>How long should the mother breast feed her child?</Text>
      <Answers>
        <Answer rank="1">002:233</Answer>
      </Answers>
      <Topic>Code of Conduct:Breastfeeding</Topic>
      <SourceContext>
        <Source>     http://forums.understanding-islam.com/showthread.php?10219-How-long-
should-the-mother-breast-feed-her-child</Source>
        <DateTaken>2012-12-11</DateTaken>
        <QuestionDate>2010-10-03</QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="27" enabled="true">
      <Text>Do men and women have equal status in Islam in respect of their deeds?</Text>
      <Answers>
        <Answer rank="1">004:124</Answer>
        <Answer rank="2">016:097</Answer>
        <Answer rank="3">033:035</Answer>
        <Answer rank="4">040:040</Answer>
        <Answer rank="5">057:012</Answer>
        <Answer rank="6">057:018</Answer>
      </Answers>
      <Topic>Islamic Ruling:Equality</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=128</Source>
        <DateTaken>2013-01-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="28" enabled="true">
      <Text>I am having a problem with jinn. I have been able to perceive jinn in
diferent ways all my life. I never seemed to be bothered by this until recently. I saw a
jinn in our apartment in the first days we moved in. Off and on I would pick up on a
little activity from this jinn or even jinns, like doors opening by themselves، seeing
it، hearing them etc... However، things seem to be changing a lot. Now something is
happening every day and things are making me uncomfortable in my home to the point where
I do not want to live here anymore. The jinn(s) open doors، shout my name startling me
out of sleep، knock over objects، appear as a cat، are messing with my computer and
phones، I see its shadows and even more. It is very freaky. I am really not sure what to
do about this problem. I am hoping moving out of this house will help the situation...?
But in the meantime I have tried to recite Al-Baqarah، Al-Ikhlas، Al-Falaq، An Nas and
even play a recording of the recitation in my house. The activity stops when I do so but
as soon as I stop the recording the jinn makes its presence known in some way or other
(most of the time). Sometimes، my recording will be shut off and my computer resets in
the middle of the recitation...which apparently has happened too many times. Jinn even
appear to me in dreams very often. I do not know how to end this situation but I would
seriously appreciste answers or suggestions soon.</Text>
      <Answers>
        <Answer rank="1">007:027</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/10513</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="29" enabled="true">
      <Text> My friend told me that Iblees (shaytan) used to be an angel. My wife says
that this is not true. Could you please give me some information?</Text>
```

```
      <Answers>
        <Answer rank="1">018:050</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/6485</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="30" enabled="true">
      <Text>In a talk recently, the speaker said that it is not fard for women to perform
Pilgrimage whereas it is only fard for men. Please could you state any ahadith or ayaah
as evidence for this statement. If this statement is wrong, please could you also state
the relevant ahadith/ayaah.</Text>
      <Answers>
        <Answer rank="1">003:097</Answer>
      </Answers>
      <Topic>Rituals:Pilgrimage</Topic>
      <SourceContext>
        <Source> http://islamqa.info/en/ref/20045</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="31" enabled="true">
      <Text>Which is the begining of the day- is it fajr or magrib. Please tell me. Which
is the middle prayer. If it is asr then, fajr is the beginning of the day. Please clarify
me in detail from the quran and the hadeeth.</Text>
      <Answers>
        <Answer rank="1">002:238</Answer>
      </Answers>
      <Topic>Rituals:Prayer</Topic>
      <SourceContext>
        <Source> http://islamqa.info/en/ref/20788</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="32" enabled="true">
      <Text>Is it permissible to pray the Isha prayer just before we go to sleep, instead
of the appointed time? Show evidence please.</Text>
      <Answers>
        <Answer rank="1">004:103</Answer>
      </Answers>
      <Topic>Rituals:Prayer</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/20788</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="33" enabled="true">
      <Text> If one finds a hole in their prayer garment, and realizes that he/she has
been praying in such a condition (the 'awra could be seen through the hole) for some
time, are all of his/her prayers in need of being repeated? If so, and the person does
not know for how long this occurred, how can they be sure they made up enough
prayers?</Text>
      <Answers>
        <Answer rank="1">007:031</Answer>
```

```
      </Answers>
      <Topic>Rituals:Prayer</Topic>
      <SourceContext>
        <Source> http://islamqa.info/en/ref/3075</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="34" enabled="true">
      <Text> I am a tuberculosis patient. I was under treatment for two years and the
doctors advised me not to fast, and scared me that if I fast the disease will come back
to me. They advised me not to fast for five years. What is the ruling on doing this?.
</Text>
      <Answers>
        <Answer rank="1">002:185</Answer>
      </Answers>
      <Topic>Rituals:Fasting</Topic>
      <SourceContext>
        <Source> http://islamqa.info/en/ref/106466</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="35" enabled="true">
      <Text> A traveller knows that he will come back tomorrow – is it permissible for
him not to fast? </Text>
      <Answers>
        <Answer rank="1">002:185</Answer>
      </Answers>
      <Topic>Rituals:Fasting</Topic>
      <SourceContext>
        <Source> http://islamqa.info/en/ref/36721</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="36" enabled="true">
      <Text> I have children and I often swear an oath to them that they should not do
such and such, but they do not listen to me. Do I have to offer any expiation in this
case?</Text>
      <Answers>
        <Answer rank="1">005:089</Answer>
      </Answers>
      <Topic>Social Norms:Oath</Topic>
      <SourceContext>
        <Source> http://islamqa.info/en/ref/127982</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="37" enabled="true">
      <Text> Someone left a sum of money with me as a trust, on the basis that he would
come back within a short period of time and collect it. But he has not got in touch and
he has not come for a long time. I called him on the number he left with me but I did not
find him there. My question is: Is it permissible for me to use this money for trade, on
the basis that I will protect it and repay him when he asks for it, and will bear in full
any loss that happens to this money?. </Text>
      <Answers>
        <Answer rank="1">004:058</Answer>
      </Answers>
```

```
      <Topic>Social Norms:Trust</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/12732</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="38" enabled="true">
      <Text> Is making Will recommended by the Quran? How many witnesses are required to
testify to the Will? What is the procedure for obtaining witness when making the will is
necessitated during journey?</Text>
      <Answers>
        <Answer rank="1">005:106</Answer>
      </Answers>
      <Topic>Code of Conduct:Will</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=290</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="39" enabled="true">
      <Text> Shall we be accounted for how did we handle trusts in our custody
irrespective of the fact whether the trust belongs to a Muslim or Non-Muslim? Is it
equally disallowed to misappropriate trust of even non believers?</Text>
      <Answers>
        <Answer rank="1">003:075</Answer>
        <Answer rank="2">003:076</Answer>
        <Answer rank="2">004:058</Answer>
      </Answers>
      <Topic>Social Norms:Trust</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=282</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="40" enabled="true">
      <Text> What are the rules for witness for men and women in case of writing down of
contacts for future dealings? Is it mandatory to provide witness when summoned to do
so?</Text>
      <Answers>
        <Answer rank="1">002:282</Answer>
      </Answers>
      <Topic>Islamic Ruling:Contract</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=282</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="41" enabled="true">
      <Text> Has every community been sent a messenger speaking the language of its
people?</Text>
      <Answers>
        <Answer rank="1">014:008</Answer>
      </Answers>
      <Topic>History:Prophets</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=131</Source>
```

```
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="42" enabled="true">
      <Text> When does the Quran specifically exhort humans to abstain from crude, lewd
speech, wicked conduct, and quarreling? </Text>
      <Answers>
        <Answer rank="1">002:197</Answer>
      </Answers>
      <Topic>Rituals:Pilgrimage</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=86</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="43" enabled="true">
      <Text> What are the criteria of devoutness according to the Quran? </Text>
      <Answers>
        <Answer rank="1">002:177</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=47</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="44" enabled="true">
      <Text> What harm can Satan inflict on believers by means of liquor and
gambling?</Text>
      <Answers>
        <Answer rank="1">005:091</Answer>
      </Answers>
      <Topic>Islamic Ruling:Haram</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=111</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="45" enabled="true">
      <Text> In how many days did Allah create the earth?</Text>
      <Answers>
        <Answer rank="1">041:009</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=326</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="46" enabled="true">
      <Text> What is the law of retribution and Qisas? </Text>
      <Answers>
        <Answer rank="1">002:178</Answer>
      </Answers>
```

```
      <Topic>Islamic Ruling:Qisas</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=102</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="47" enabled="true">
      <Text> What is the procedure of payment of dower money if the divorce takes place
before the marriage is consummated and dower money had been fixed? </Text>
      <Answers>
        <Answer rank="1">002:237</Answer>
      </Answers>
      <Topic>Code of Conduct:Dowry</Topic>
      <SourceContext>
        <Source>http://studythequran.com/ShowQandA.php?qa_id=63</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="48" enabled="true">
      <Text>Longtime ago some body told me the fard of wudu is not to wash your arms, is
to wash your mouth. Since then I have not beang washing my arms in wudu, now yesterday
somebody told me washing your mouth is not fard in wudu, washing your arms is fard in
wudu. Now do I have to repeat all my salahs because I didnt know about the arms?</Text>
      <Answers>
        <Answer rank="1">005:006</Answer>
      </Answers>
      <Topic>Islamic Ruling:Ablution</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/cat/141/ref/islamqa/198415</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="49" enabled="true">
      <Text>What is the ruling on drinking beer and similar alcoholic drinks?</Text>
      <Answers>
        <Answer rank="1">005:090</Answer>
        <Answer rank="2">005:091</Answer>
      </Answers>
      <Topic>Islamic Ruling:Haram</Topic>
      <SourceContext>
        <Source>http://islamqa.com/en/cat/145/ref/islamqa/127020</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="50" enabled="true">
      <Text>Is the doctors' knowledge of the fetus in the uterus complete?</Text>
      <Answers>
        <Answer rank="1">031:034</Answer>
        <Answer rank="2">013:008</Answer>
        <Answer rank="3">053:032</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source>http://islamqa.com/en/ref/1513</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
```

```
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="51" enabled="true">
      <Text>When I was a kid I was told that when God sent Iblis out of the Heavens and
looking at the rage of Allah, the Angels bowed a second time thus the reason for us to
Bow twice in our prayers. Is there any truth in it ? as I can't find any references.
Could you please clarify ?</Text>
      <Answers>
        <Answer rank="1">002:034</Answer>
        <Answer rank="1">018:050</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source>http://islamqa.com/en/ref/1513</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="52" enabled="true">
      <Text>As Salamu Alakium, All my life i lived in America and I know what was right
and what was wrong enough to lead my life there. However now I am back to my country and
there lot of things that people do that is not according to Islam. I really need your
help on who is my mahram and who can I go in front of. Can I go in front of my mother's
uncle amd her male cousins? I mean I can't marry them so why can't I go in front of
them.Thank you for answering my question.</Text>
      <Answers>
        <Answer rank="1">004:022</Answer>
        <Answer rank="2">004:023</Answer>
      </Answers>
      <Topic>Islamic Ruling:Unmarriageable Kin</Topic>
      <SourceContext>
        <Source><![CDATA[http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=5583&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2013-05-02</DateTaken>
        <QuestionDate>2005-03-15</QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="53" enabled="true">
      <Text>Being the bride's father my father on prompt from the guy's family was the
first one to ask for the mahr. He had consulted me in this regard before hand, what I had
wanted was coming out more then normally people over the years have asked in our family.
After consultation with my father I agreed upon letting him ask for what was the norm
within our family. My father then was the first one to say the amount or "price" as
people may call it. My first question to you is..is it wrong for the bride or bride's
family to first say an amount? Is the fact that the father actually says an amount an
indication that he is trying to sell his daughter or put a price on her??? Secondly is
there a minimum or maximum that the bride can ask for? I am asking these questions as
these have been highlighted to me by the groom and I would like to clearify whether this
action was right or wrong on mine and my father's part.Please do guide me. thanks.</Text>
      <Answers>
        <Answer rank="1">004:004</Answer>
        <Answer rank="2">004:024</Answer>
      </Answers>
      <Topic>Code of Conduct:Mahr</Topic>
      <SourceContext>
        <Source><![CDATA[     http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=6487&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2013-05-02</DateTaken>
        <QuestionDate>2005-05-02</QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
```

```
    </Question>
    <Question id="54" enabled="true">
      <Text>It is possible for a woman to demand divorce, provided that she repays her
dower or meher, to her husband. What if she does not have the money to repay? Does this
mean that she cannot be divorced? In case the divorce is granted, who has the custody of
the children? </Text>
      <Answers>
        <Answer rank="1">002:229</Answer>
      </Answers>
      <Topic>Code of Conduct:Divorce</Topic>
      <SourceContext>
        <Source><![CDATA[     http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=4546&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2013-05-02</DateTaken>
        <QuestionDate>1995-01-01</QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="55" enabled="true">
      <Text> I made a promise to Allah that if He did something for me, I would do
something else in return (something specific). (This is called a Nedr in arabic).Well, He
granted my request, and now I have to live up to my promise but it is just too hard for
me. I want to please Him but am too weak to do what I promised. Is there a way out
without incurring the anger of Allah? What can I do?</Text>
      <Answers>
        <Answer rank="1">005:089</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source><![CDATA[     http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=5725&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2013-05-02</DateTaken>
        <QuestionDate> 2005-03-23 </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="56" enabled="true">
      <Text> please tell me the do's and dont's of preforming gusl before sahour...if a
couple have mated and do not have time to preform ghusal...can they eat or drink in that
state?...is yes...can they pray fajar after eating and then preforming ghusal? please
advise. </Text>
      <Answers>
        <Answer rank="1">002:187</Answer>
      </Answers>
      <Topic>Rituals:Fasting</Topic>
      <SourceContext>
        <Source><![CDATA[     http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=28831&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2013-05-02</DateTaken>
        <QuestionDate> 2008-09-10 </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="57" enabled="true">
      <Text> Assalamuallaikum, Could you please tell me verses in Qur'an and hadith about
forgiving others? </Text>
      <Answers>
        <Answer rank="1">002:109</Answer>
        <Answer rank="2">005:013</Answer>
        <Answer rank="3">024:022</Answer>
        <Answer rank="4">043:089</Answer>
        <Answer rank="5">064:014</Answer>
      </Answers>
      <Topic>Social Norms:Forgiving</Topic>
```

```
      <SourceContext>
        <Source><![CDATA[     http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=1886&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2013-05-02</DateTaken>
        <QuestionDate> 1997-11-03 </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="58" enabled="true">
      <Text>Is all sea food lawful?</Text>
      <Answers>
        <Answer rank="1">005:096</Answer>
      </Answers>
      <Topic>Islamic Ruling:Food</Topic>
      <SourceContext>
        <Source>http://www.studythequran.com/ShowQandA.php?qa_id=71</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="59" enabled="true">
      <Text>I brought in some foreign workers about eight years ago, and sold the visas
to another person here, on the basis that he would bring in the workers and they would do
whatever work he wanted, i.e., they would not be working with me in my company. This was
a condition between me and them, and we agreed that they would pay a portion of their
salary to me at the end of each month, but this was not binding. Some of the workers have
gone back to their own country, and I have no address for them, and some of them are
still here, but I do not take any money from them. What is the ruling on the money which
I have taken for selling the visas and the money I have taken from the workers?</Text>
      <Answers>
        <Answer rank="1">002:188</Answer>
      </Answers>
      <Topic>Islamic Ruling:Business</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/9385/</Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="60" enabled="true">
      <Text> Is it halal, haram, or makruh to eat non zabiha meat(beef)? </Text>
      <Answers>
        <Answer rank="1">005:003</Answer>
      </Answers>
      <Topic>Islamic Ruling:Food</Topic>
      <SourceContext>
        <Source><![CDATA[http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=979&-format=detailpop.shtml&-find]]></Source>
        <DateTaken>2013-05-02</DateTaken>
        <QuestionDate> 1997-04-08 </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="61" enabled="true">
      <Text> A'Salaam O'lekum Is it fine for us, muslims, to eat animals that have been
shot with a bullet. A moose, for example, could be shot in the lungs so that the blood
would rush to that area, leaving the meat clean. A moose would also be hung for a few
days. Therefore, measures would be taken to draw out the blood. Thank you for listening.
</Text>
      <Answers>
        <Answer rank="1">005:003</Answer>
      </Answers>
```

```
      <Topic>Islamic Ruling:food</Topic>
      <SourceContext>
        <Source><![CDATA[http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=2071&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2013-05-02</DateTaken>
        <QuestionDate> 1997-11-22 </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="62" enabled="true">
      <Text>I have make the big sin adultery but not with the same man for 4 years, and I
am very sincere to stop. I want to change my life , i want my heart to follow allah
direction . i feel my soul sick and i can't continues to this sin anymore. i want allah
to forgive me and i am sincère to pray and change my life . How can i know allah accept
my taoba ? the last man i have commit the sin, want to marry me but he dont pray even he
dont want to stop the sin till we get marry. what can i do ? do i must leave him ?</Text>
      <Answers>
        <Answer rank="1">024:003</Answer>
        <Answer rank="1">004:017</Answer>
        <Answer rank="1">004:018</Answer>
      </Answers>
      <Topic>Islamic Ruling:Adultery</Topic>
      <SourceContext>
        <Source><![CDATA[     http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=26314&-format=detailpop.shtml&-find]]></Source>
        <DateTaken>2013-05-02</DateTaken>
        <QuestionDate>2008-02-19</QuestionDate>
        <QuestionType> Derived User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="63" enabled="true">
      <Text>What is the Fatwa about a person who intentionaly says to his wife that she
is his mother? Is it Zihar or not?</Text>
      <Answers>
        <Answer rank="1">058:002</Answer>
      </Answers>
      <Topic>Islamic Ruling:Zihar</Topic>
      <SourceContext>
        <Source><![CDATA[
http://www.islamweb.net/emainpage/index.php?page=showfatwa&Option=FatwaId&Id=85470
]]></Source>
        <DateTaken>2013-01-16</DateTaken>
        <QuestionDate>2003-01-22</QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="64" enabled="true">
      <Text>Salam, my friend commited zihar by refering to his wife as his mothers
back.He want to get her back by perfoming kafarah,however he sees it difficult to fast
two consegative months and instead want to feed sixty poor people because he has the
money.is this allowed since he is capable of fasting since he is still young and
energetic.</Text>
      <Answers>
        <Answer rank="1">058:003</Answer>
      </Answers>
      <Topic>Islamic Ruling:Zihar</Topic>
      <SourceContext>
        <Source><![CDATA[
http://www.islamweb.net/emainpage/index.php?page=showfatwa&Option=FatwaId&Id=151039
]]></Source>
        <DateTaken>2013-01-16</DateTaken>
        <QuestionDate>2011-03-13</QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
```

```
    </Question>
    <Question id="65" enabled="true">
      <Text>Respected scholars, as-salamu `alaykum. I got married, and before the wedding
ceremony, I slept with my husband, but I am still virgin, and it was more than once, but
he was careful that I keep my virginity before the ceremony. We fought, and he divorced
me. He went away to get married from his country. Do I need to wait for three months
before I can get married again (knowing that I am still virgin)? And in case he comes
back, how should we get together again? Thank you.</Text>
      <Answers>
        <Answer rank="1">033:049</Answer>
      </Answers>
      <Topic>Code of Conduct:Waiting Period</Topic>
      <SourceContext>
        <Source><![CDATA[  http://www.onislam.net/english/ask-the-scholar/family/waiting-
period/176502.html ]]></Source>
        <DateTaken>2013-01-24</DateTaken>
        <QuestionDate>2008-03-26</QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="66" enabled="true">
      <Text>Brother X and Sister Y were married for a short while before things turned
sour. The couple never lived together and the marriage was never consummated (even though
they had the opportunity too). The couple have been separated for several months (more
than that of a waiting period) and the divorce has not been finalized yet. The whole
process (for divorce) has been initiated by Sister Y. If the marriage was not consummated
then do the normal divorce rulings apply or is the marriage annulled (as opposed to
divorced)?</Text>
      <Answers>
        <Answer rank="1">033:049</Answer>
      </Answers>
      <Topic>Code of Conduct:Waiting Period</Topic>
      <SourceContext>
        <Source><![CDATA[        http://spa.qibla.com/issue_view.asp?HD=12&ID=173&CATE=11
]]></Source>
        <DateTaken>2013-01-24</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="67" enabled="true">
      <Text>Should the debt agreements be documented and what are the rules for
documenting fixed-term debt?</Text>
      <Answers>
        <Answer rank="1">002:282</Answer>
      </Answers>
      <Topic>Islamic Ruling:Debt</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=40 ]]></Source>
        <DateTaken>2013-01-24</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="68" enabled="true">
      <Text>Is it enjoined on believers to look after the orphans, maintain them, take
care of their property and hand over their property to them after they are grown up?
</Text>
      <Answers>
        <Answer rank="1">002:219</Answer>
        <Answer rank="1">002:220</Answer>
        <Answer rank="1">006:152</Answer>
      </Answers>
      <Topic>Code of Conduct:Orphans</Topic>
```

```
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=159 ]]></Source>
        <DateTaken>2013-01-24</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="69" enabled="true">
      <Text>What sort of volume I should recite Quran in when offering my Namaz?</Text>
      <Answers>
        <Answer rank="1">017:110</Answer>
      </Answers>
      <Topic>Rituals:Prayer</Topic>
      <SourceContext>
        <Source><![CDATA[ Offline ]]></Source>
        <DateTaken>2012-07-14</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Offline User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="70" enabled="true">
      <Text>My father has been on dailysis for about 3years,as required of those patients
not to eat foods high in phosphate and pottasium but he becomes abusive when i don't let
him eat those items,In return i have to be a little rude and disobedient and not give
respect to him and literally fight with to keep him away from foods that may cause
serious problem i don't like myself to be that way but that atleast keeps him away from
things that may hurt him.What Quran says about it </Text>
      <Answers>
        <Answer rank="1">017:023</Answer>
        <Answer rank="1">017:024</Answer>
        <Answer rank="1">017:025</Answer>
      </Answers>
      <Topic>Code of Conduct:Family</Topic>
      <SourceContext>
        <Source><![CDATA[     http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=3300&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2012-07-14</DateTaken>
        <QuestionDate> 1998-06-07 </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="71" enabled="true">
      <Text>Is there veiling for men also?</Text>
      <Answers>
        <Answer rank="1">024:030</Answer>
      </Answers>
      <Topic>Islamic Ruling:Veiling</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=288 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="72" enabled="true">
      <Text> He died and he had a wife, a son and a daughter, and a mother and a father.
How should the house and car that are inherited be divided?</Text>
      <Answers>
        <Answer rank="1">004:011</Answer>
        <Answer rank="2">004:012</Answer>
      </Answers>
      <Topic>Code of Conduct:Inheritance</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/171811/inheritance</Source>
```

```
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="73" enabled="true">
      <Text> Why are adultery, eating the flesh of swine forbidden in Islam or consedered
"haram"?</Text>
      <Answers>
        <Answer rank="1">017:032</Answer>
        <Answer rank="1">002:173</Answer>
      </Answers>
      <Topic>Islamic Ruling:Haram</Topic>
      <SourceContext>
        <Source> http://islamqa.info/en/ref/751</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="74" enabled="true">
      <Text> What is the wisdom behind the Islamic prohibition on gambling?</Text>
      <Answers>
        <Answer rank="1">005:090</Answer>
        <Answer rank="1">005:091</Answer>
      </Answers>
      <Topic>Islamic Ruling:Haram</Topic>
      <SourceContext>
        <Source>http://islamqa.info/en/ref/4013</Source>
        <DateTaken>2013-03-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="75" enabled="true">
      <Text>Can Isa(as), Musa(as), and other Prophets be referred to as “Muslims”?</Text>
      <Answers>
        <Answer rank="1">003:067</Answer>
        <Answer rank="2">010:084</Answer>
        <Answer rank="3">005:111</Answer>
      </Answers>
      <Topic>History:Prophets</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=134 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="76" enabled="true">
      <Text>Are there such types of people who misled others by their good words but in
fact spread corruption on earth and are engaged in destruction? (Like many political
leaders of different communities). What will be the end of such people?</Text>
      <Answers>
        <Answer rank="1">002:204</Answer>
        <Answer rank="2">002:205</Answer>
        <Answer rank="3">002:206</Answer>
      </Answers>
      <Topic>Islamic Ruling:Corruption</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=30 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
```

```
      </SourceContext>
    </Question>
    <Question id="77" enabled="true">
      <Text>When does the Quran specifically exhort humans to abstain from crude, lewd
speech, wicked conduct, and quarreling?</Text>
      <Answers>
        <Answer rank="1">002:197</Answer>
      </Answers>
      <Topic>Rituals:Pilgrimage</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=86 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="78" enabled="true">
      <Text>A family of 3 brothers, 2 is very rich and one of them performs Hajj every
year with his family of 5 dependents for last 3 years and spent millions of Pakistani
Rupees every year on Hajj. The 3rd brother is poor. Q: What Islam says about performing
Hajj many times even his own brother are very poor?</Text>
      <Answers>
        <Answer rank="1">024:022</Answer>
        <Answer rank="1">051:019</Answer>
      </Answers>
      <Topic>Rituals:Pilgrimage</Topic>
      <SourceContext>
        <Source><![CDATA[     http://www.islamicity.com/qa/action.lasso.asp?-db=services&-
lay=Ask&-op=eq&number=33311&-format=detailpop.shtml&-find ]]></Source>
        <DateTaken>2012-07-14</DateTaken>
        <QuestionDate> 2009-08-11</QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="79" enabled="true">
      <Text>What should pilgrims do, if they cannot afford a sacrifice at the time of the
pilgrimage?</Text>
      <Answers>
        <Answer rank="1">002:196</Answer>
      </Answers>
      <Topic>Rituals:Pilgrimage</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=85 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="80" enabled="true">
      <Text>how to wear ahram and what are the things to avoid while i am in
ihram?</Text>
      <Answers>
        <Answer rank="1">002:196</Answer>
        <Answer rank="2">002:197</Answer>
        <Answer rank="3">005:095</Answer>
      </Answers>
      <Topic>Rituals:Pilgrimage</Topic>
      <SourceContext>
        <Source><![CDATA[ Offline ]]></Source>
        <DateTaken>2012-09-18</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Offline User Query </QuestionType>
      </SourceContext>
    </Question>
```

```
    <Question id="81" enabled="true">
      <Text>Can an adopted son of a man become his real son?</Text>
      <Answers>
        <Answer rank="1">033:004</Answer>
      </Answers>
      <Topic>Code of Conduct:Adoption</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=352 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="82" enabled="true">
      <Text>A general belief is in vogue among majority of Muslims that whatever they do
they will be forgiven and will enter paradise because they belong to a privileged group.
What does the Quran say about such belief?</Text>
      <Answers>
        <Answer rank="1">007:169</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=299 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="83" enabled="true">
      <Text>What relationship do the wives of the prophet hold with the Muslim people?
</Text>
      <Answers>
        <Answer rank="1">033:006</Answer>
      </Answers>
      <Topic>History:Prophets</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=367 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="84" enabled="true">
      <Text>Is every hardship followed by relief?</Text>
      <Answers>
        <Answer rank="1">094:005</Answer>
        <Answer rank="1">094:006</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=386 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="85" enabled="true">
      <Text>What are jinn made of?</Text>
      <Answers>
        <Answer rank="1">055:015</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=318 ]]></Source>
```

```
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="86" enabled="true">
      <Text>What is the condition for marriage with a slave woman?</Text>
      <Answers>
        <Answer rank="1">004:025</Answer>
      </Answers>
      <Topic>Code of Condcut:Slaves</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=317 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="87" enabled="true">
      <Text>How near is Allah to us?</Text>
      <Answers>
        <Answer rank="1">050:016</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=319 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="88" enabled="true">
      <Text>have the following question: Is it allowed for a small child to see its
parents' 'awrah' (private parts which are required to be covered in public)? Is it for
example allowed to shower together with its parents? </Text>
      <Answers>
        <Answer rank="1">024:058</Answer>
      </Answers>
      <Topic>Code of Condcut:Privacy</Topic>
      <SourceContext>
        <Source><![CDATA[ http://islamqa.info/en/ref/102187 ]]></Source>
        <DateTaken>2012-08-16</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="89" enabled="true">
      <Text>With regards to greeting salaam. I have notice some muslim people don't greet
salaam when they arrive so it prompted me to greet them salaam which is the opposite of
what should be. My question is am I doing bidaa because of this (greeting salaam to the
person who arrives who does not say salaam).. did I sin and second thing should I greet
next time with salaam a person who just arrive with he / she didn't say salaam. </Text>
      <Answers>
        <Answer rank="1">024:061</Answer>
      </Answers>
      <Topic>Social Norms:Manners</Topic>
      <SourceContext>
        <Source><![CDATA[ http://islamqa.info/en/ref/72231/ ]]></Source>
        <DateTaken>2012-08-19</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="90" enabled="true">
```

```
      <Text>In which Lunar month the Quran was revealed?</Text>
      <Answers>
        <Answer rank="1">002:185</Answer>
      </Answers>
      <Topic>Quran</Topic>
      <SourceContext>
        <Source><![CDATA[ Offline ]]></Source>
        <DateTaken>2012-10-29</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Offline User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="91" enabled="true">
      <Text> In North Pole and South Pole sun never sets for six month and never rises
for six months. The sun rises from east and when it is reaches the west it does not set
but it rises again, this process continues for six months. How can we explain/understand
the hadeeth that says 'sun will rise from west before judgment day'?. </Text>
      <Answers>
        <Answer rank="1">006:158</Answer>
        <Answer rank="2">036:038</Answer>
        <Answer rank="3">039:005</Answer>
        <Answer rank="4">021:033</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source><![CDATA[http://islamqa.info/en/ref/71201 ]]></Source>
        <DateTaken>2012-08-19</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="92" enabled="true">
      <Text>I have a fiancee who is giving me problems with money. She wants so many
expensive things just for the wedding and the honeymoon. I keep saying that we shouldnt
spend too much money on these things because it will lose baraka, but she just says its a
once in a lifetime thing. Is she right? Is spending alot of money just for a wedding and
a one week honeymoon justified just because it's only once in our life? Please let me
know, and I already know about spending on the wife is good, but I'm not referring to
food and clothing and housing, only the extravagant things. </Text>
      <Answers>
        <Answer rank="1">025:067</Answer>
        <Answer rank="2">007:031</Answer>
      </Answers>
      <Topic>Code of Conduct:Extravagance</Topic>
      <SourceContext>
        <Source><![CDATA[http://islamqa.info/en/ref/171265 ]]></Source>
        <DateTaken>2012-08-19</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="93" enabled="true">
      <Text>It is not possible to see the new moon with the naked eye before it is 30
hours old, and in addition to that, it is sometimes not possible to see it at all because
of the weather conditions. On this basis, is it permissible to resort to using
astronomical information to calculate the likely time for sighting the new moon and the
start of Ramadaan, or do we have to actually sight the new moon before we start fasting
the blessed month of Ramadaan?</Text>
      <Answers>
        <Answer rank="1">002:185</Answer>
        <Answer rank="2">002:189</Answer>
      </Answers>
      <Topic>Rituals:Fasting</Topic>
      <SourceContext>
```

```
        <Source><![CDATA[ http://islamqa.info/en/ref/1245 ]]></Source>
        <DateTaken>2012-08-19</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="94" enabled="true">
      <Text> I am 24 years old and got married in 2004. I have been trying to conceive
since this time and only recently been successful, however, my husband was not happy
about this and has decided to leave me because he doesn't want to afford a baby. When he
asked me to have an abortion and choose between him and the baby, I chose to keep the
baby. So my first question is that was this Islamically the correct choice to make?
</Text>
      <Answers>
        <Answer rank="1">017:031</Answer>
        <Answer rank="2">006:151</Answer>
      </Answers>
      <Topic>Islamic Ruling:Abortion</Topic>
      <SourceContext>
        <Source><![CDATA[ Offline ]]></Source>
        <DateTaken>2012-10-29</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Offline User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="95" enabled="true">
      <Text>Sometimes when I'm in the market place and I am in a store and have chosen
something and am at the checkout, the Azan (call to prayer) is given. Is it permissible
for me to complete the purchase or is that haraam? </Text>
      <Answers>
        <Answer rank="1">062:009</Answer>
        <Answer rank="1">062:010</Answer>
      </Answers>
      <Topic>Rituals:Prayer</Topic>
      <SourceContext>
        <Source><![CDATA[ http://islamqa.info/en/ref/140662 ]]></Source>
        <DateTaken>2012-08-16</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="96" enabled="true">
      <Text> We hastened to leave Hajj early and departed from Mina before sunset, so
that we would not have to stay until the third day. After leaving Mina we came back and
stoned the jamaraat after sunset. Is what we did permissible? </Text>
      <Answers>
        <Answer rank="1">002:203</Answer>
      </Answers>
      <Topic>Rituals:Pilgrimage</Topic>
      <SourceContext>
        <Source><![CDATA[ http://islamqa.info/en/ref/43274 ]]></Source>
        <DateTaken>2012-10-26</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="97" enabled="true">
      <Text> Is it permissible for an elderly woman to uncover her face before men who
are not her mahrams?. </Text>
      <Answers>
        <Answer rank="1">024:060</Answer>
      </Answers>
      <Topic>Islamic Ruling:Veiling</Topic>
      <SourceContext>
```

```
        <Source><![CDATA[http://islamqa.info/en/ref/111940/ ]]></Source>
        <DateTaken>2012-10-26</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Online User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="98" enabled="true">
      <Text> Can only the believers maintain the mosque? </Text>
      <Answers>
        <Answer rank="1">009:018</Answer>
      </Answers>
      <Topic>Code of Conduct:Mosque</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=64 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> FAQ </QuestionType>
      </SourceContext>
    </Question>
    <Question id="99" enabled="true">
      <Text> In what orders seven heavens have been created by Allah?</Text>
      <Answers>
        <Answer rank="1">067:003</Answer>
        <Answer rank="1">071:015</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source><![CDATA[ http://studythequran.com/ShowQandA.php?qa_id=364 ]]></Source>
        <DateTaken>2012-11-22</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Derived User Query </QuestionType>
      </SourceContext>
    </Question>
    <Question id="100" enabled="true">
      <Text> I have been told that there are angels who record our deeds on daily basis.
How true is this? </Text>
      <Answers>
        <Answer rank="1">050:017</Answer>
      </Answers>
      <Topic>Faith:Belief</Topic>
      <SourceContext>
        <Source><![CDATA[ Offline ]]></Source>
        <DateTaken>2012-10-26</DateTaken>
        <QuestionDate> Unknown </QuestionDate>
        <QuestionType> Offline User Query </QuestionType>
      </SourceContext>
    </Question>
  </Questions>
</QuestionSet>
```

Appendix 4 – PARMS Output

Answer Analysis Report

A	B	C	D
Question Id	**Answer Found**	**Question Text**	**Question Terms**
01	No	Assalamualaikum.. Is crab halal or haram so as shrimps am confused...Thanks	assalamualaikum,crab,haram,confusedthank,halal,shrimp
Answer Id	**Weight**	**Answer Text**	**Answer Terms**
Answer Id	**Location**	**Answer Text**	**Answer Terms**
005:096	0 / 0 / 6236	Lawful to you is the game of the sea and its food, a provision for you and for the travellers, and the game of the land is forbidden to you so long as you are on pilgrimage, and be careful of (your duty to) Allah, to Whom you shall be gathered.	pilgrimag,travel,care,long,land,gather,duti,food,law,provis,sea,forbidden,game
Question Id	**Answer Found**	**Question Text**	**Question Terms**
02	No	I am the head of a household, with a wife and two children. My salary barely covers the necessities of life, and I do not have any other income. I worked in one of the Gulf countries for 4 years, and I saved an amount of money which I deposited in one of the Islamic banks to provide me with an income to help me deal with different burdens of life. My salary and this income from this investment are just sufficient for me and my family. Am I obliged to spend some of this deposited money on Hajj expenses, and am I obliged to go for Hajj in the light of these circumstances? Please note that if I spend this amount from my bank account on Hajj expenses, this will affect my monthly income and will make things very difficult for me financially. What do you advise me to do?	financi,month,suffici,invest,year,amount,account,islam,provid,head,famili,necess,wife,advis,children,note,save,gulf,circumst,4,expens,light,deposit,make,household,deal,incom,salari,difficult,cover,burden,work,bare,hajj,money,affect,thing,life,spend,countri,bank,oblig
Answer Id	**Weight**	**Answer Text**	**Answer Terms**
007:189	25.44955044	He it is Who created you from a single being, and of the same (kind) did He make his mate, that he might incline to her; so when he covers her she bears a light burden, then moves about with it; but when it grows heavy, they both call upon Allah, their Lord: If Thou givest us a good one, we shall certainly be of the grateful ones.	bear,thou,lord,make,light,singl,mate,kind,good,burden,creat,givest,grate,inclin,move,grow,heavi,cover,call
002:233	23.97222762	And the mothers should suckle their children for two whole years for him who desires to make complete the time of suckling; and their maintenance and their clothing must be-- borne by the father according to usage; no soul shall have imposed upon it a duty but to the extent of its capacity; neither shall a mother be made to suffer harm on account of her child, nor a father on account of his child, and a similar duty (devolves) on the (father's) heir, but if both desire weaning by mutual consent and counsel, there is no blame	born,suckl,soul,capac,promis,devolv,extent,pay,harm,year,care,child,account,mainten,suffer,counsel,mother,desir,wean,children,similar,cloth,long,mutual,father,heir,make,duti,impos,blame,see,made,consent,time,engag,wetnurs,usag,complet

Results Summary Report

A	B	C	D	E	F	G	H
Translated By	**Remove Stopwords**	**Stem Words**	**Stemmer**	**Weight Function**	**Results Cut-off Point**	**Term Expansion Level**	**Ignore Duplicates**
Shakir	Yes	Yes	Snowball	idf	10	None	No
	Count	**Percentage**	**Ids**				
Total Questions	100						
Success Results	34/98	34.69	03,06,08,11,12,14,19,20,24,25,26,29,31,36,40,47,48,51,53,56,58,61,65,67,68,72,73,81,82,83,89,90,93,98				
Failed Results	66/98		01,02,04,05,07,09,10,13,15,16,17,18,21,22,23,27,28,30,32,33,34,35,37,38,39,41,42,43,44,45,46,49,50,52,54,55,57,59,60,62,63,64,(				
No Results	2						

01	**02**	**03**	**04**	**05**	**06**	**07**	**08**
0	0	1	0	0	1	0	1

Question Id	**Answer Found**
01	No
02	No
03	Yes
04	No
05	No
06	Yes
07	No
08	Yes
09	No
10	No
11	Yes
12	Yes
13	No
14	Yes
15	No
16	No
17	No
18	No
19	Yes
20	Yes
21	No
22	No
23	No
24	Yes
25	Yes
26	Yes
27	No
28	No
29	Yes
30	No
31	Yes
32	No

Result Statistical Report

Total Que	Total Cor	Total Retr	Correct R	Correct R	Accuracy	Precision	Recall	MAP	MRR
100	156	98	34	36	0.34	0.346939	0.230769	0.216873	0.63366

Question: 01			
Rank	Verse Id	Relevant	Precision
Average:			
Reciprocal Rank:			

Question: 02			
Rank	Verse Id	Relevant	Precision
1	007:189		
2	002:233		
3	007:157		
4	002:259		
5	002:264		
6	002:265		
7	065:004		
8	002:196		
9	004:127		
10	010:093		
Average:			
Reciprocal Rank:			

Question: 03			
Rank	Verse Id	Relevant	Precision
1	004:011	X	1
2	004:012		
3	004:176		
4	024:061		
5	042:014		
6	033:006		
7	005:003		
8	009:044		
9	004:023		
10	007:150		
Average:			1
Reciprocal Rank:			1

Control Vocabulary

CV 2

For TestRun_Shakir_ControlVocabLevel2

#001:001
benefic1
merci 1
/2

#001:002
prais 1
world 1
due 1
lord 1
/4

#001:003
benefic1
merci 1
/2

#001:004
day 1
judgment 1
master1
/3

#001:005
beseech 1
thee 2
serv 1
/4

#001:006
path 1
/1

#001:007

```
        wrath  1
        brought         1
        astray  1
        thi     1
        thou    1
        bestow1
        hast    1
        favor   1
        path    2
/10

#002:001
        alif    1
        lam     1
        mim     1
/3

#002:002
        guid    1
        book    1
        guard   1
        evil    1
        doubt   1
/5

#002:003
        spend   1
        prayer  1
        unseen          1
/3

#002:004
        reveal  2
/2

#002:005
        lord    1
        success         1
/2
```

```
#002:006
        alik    1
        sure    1
        warn    2
        disbeliev       1
/5

#002:007
        punish 1
        eye     1
        heart   1
        hear    1
        great   1
        seal    1
        cover   1
        set     1
/8

#002:008
        day     1
        believ  1
        peopl   1
/3

#002:009
        perceiv         1
        deceiv 2
        desir   1
/4

#002:010
        heart   1
        chastis 1
        pain    1
        diseas 2
        ad      1
/6
```

#002:011
peacemak 1
land 1
make 1
mischief 1
/4

#002:012
maker 1
perceiv 1
mischief 1
sure 1
/4

#002:013
fool 2
sure 1
peopl 1
/4

#002:014
mock 1
meet 1
sure 1
shaitan 1
/4

#002:015
back 1
mockeri 1
blind 1
leav 1
wander 1
pay 1
inordinaci 1
/7

#002:016
follow 1

bargain 1
error 1
buy 1
gain 1
bring 1
direct 2
/8

#002:017
kindl 1
light 1
dark 1
fire 1
parabl 2
illumin 1
left 1
utter 1
/9

#002:018
dumb 1
back 1
deaf 1
turn 1
blind 1
/5

#002:019
encompass 1
rain 1
death 1
unbeliev 1
ear 1
peal 1
cloud 1
put 1
lightn 1
fear 1
thunder 2

utter 1
dark 1
finger 1
abund 1
/16

#002:020
shine 1
power 1
hear 1
stand 1
take 1
pleas 1
sure 1
lightn 1
dark 1
sight 2
walk 1
thing 1
/13

#002:021
creat 1
guard 1
men 1
evil 1
lord 1
serv 1
/6

CV 3

For TestRun_Shakir_ControlVocabLevel3

abandon
002:246 1 23 01.00 11.61 11.61 00.54
006:120 1 09 01.00 11.61 11.61 00.92
/2
abas
002:061 1 32 01.00 07.70 07.70 00.42
003:112 1 20 01.00 07.70 07.70 00.59
003:127 1 08 01.00 07.70 07.70 00.97
003:146 1 12 01.00 07.70 07.70 00.80
004:014 1 08 01.00 07.70 07.70 00.97
007:119 1 03 01.00 07.70 07.70 01.32
009:063 1 09 01.00 07.70 07.70 00.92
010:027 1 15 01.00 07.70 07.70 00.70
016:048 1 09 01.00 07.70 07.70 00.92
022:018 1 17 01.00 07.70 07.70 00.65
025:069 1 06 01.00 07.70 07.70 01.08
027:037 1 09 01.00 07.70 07.70 00.92
027:087 1 07 01.00 07.70 07.70 01.02
028:082 1 17 01.00 07.70 07.70 00.65
031:006 1 12 01.00 07.70 07.70 00.80
034:014 1 16 01.00 07.70 07.70 00.68
040:060 1 09 01.00 07.70 07.70 00.92
041:016 2 13 02.00 07.70 15.40 01.13
041:017 1 10 01.00 07.70 07.70 00.87
042:045 1 16 01.00 07.70 07.70 00.68
044:030 1 05 01.00 07.70 07.70 01.15
045:009 1 05 01.00 07.70 07.70 01.15
046:020 1 17 01.00 07.70 07.70 00.65
056:003 1 03 01.00 07.70 07.70 01.32
058:005 1 14 01.00 07.70 07.70 00.73
058:016 1 08 01.00 07.70 07.70 00.97
058:020 1 04 01.00 07.70 07.70 01.23
059:005 1 09 01.00 07.70 07.70 00.92
066:008 1 28 01.00 07.70 07.70 00.47
068:043 1 07 01.00 07.70 07.70 01.02

/30
abasest
003:026 1 28 01.00 12.61 12.61 00.47
/1
abat
011:044 1 15 01.00 11.61 11.61 00.70
043:075 1 02 01.00 11.61 11.61 01.42
/2
abhor
026:168 1 03 01.00 11.61 11.61 01.32
049:012 1 19 01.00 11.61 11.61 00.61
/2
abid
002:025 1 14 01.00 06.13 06.13 00.73
002:039 1 06 01.00 06.13 06.13 01.08
002:081 1 09 01.00 06.13 06.13 00.92
002:082 1 05 01.00 06.13 06.13 01.15
002:125 1 21 01.00 06.13 06.13 00.57
002:162 1 04 01.00 06.13 06.13 01.23
002:217 1 35 01.00 06.13 06.13 00.39
002:257 1 12 01.00 06.13 06.13 00.80
002:275 1 25 01.00 06.13 06.13 00.51
003:015 1 15 01.00 06.13 06.13 00.70
003:088 1 04 01.00 06.13 06.13 01.23
003:116 1 08 01.00 06.13 06.13 00.97
003:136 1 11 01.00 06.13 06.13 00.83
003:198 1 11 01.00 06.13 06.13 00.83
004:013 1 12 01.00 06.13 06.13 00.80
004:014 1 08 01.00 06.13 06.13 00.97
004:057 1 15 01.00 06.13 06.13 00.70
004:093 1 12 01.00 06.13 06.13 00.80
004:122 1 12 01.00 06.13 06.13 00.80
004:169 1 04 01.00 06.13 06.13 01.23
005:080 1 07 01.00 06.13 06.13 01.02
005:085 1 08 01.00 06.13 06.13 00.97
005:119 1 13 01.00 06.13 06.13 00.76
006:128 1 25 01.00 06.13 06.13 00.51
007:036 1 07 01.00 06.13 06.13 01.02
007:042 1 09 01.00 06.13 06.13 00.92

009:017 1 10 01.00 06.13 06.13 00.87
009:022 1 04 01.00 06.13 06.13 01.23
009:063 1 09 01.00 06.13 06.13 00.92
009:068 1 12 01.00 06.13 06.13 00.80
009:072 1 20 01.00 06.13 06.13 00.59
009:089 1 08 01.00 06.13 06.13 00.97
009:100 1 14 01.00 06.13 06.13 00.73
010:026 1 10 01.00 06.13 06.13 00.87
010:027 1 15 01.00 06.13 06.13 00.70
010:052 1 06 01.00 06.13 06.13 01.08
010:087 1 16 01.00 06.13 06.13 00.68
011:023 1 07 01.00 06.13 06.13 01.02
011:107 1 11 01.00 06.13 06.13 00.83
011:108 1 11 01.00 06.13 06.13 00.83
013:005 1 10 01.00 06.13 06.13 00.87
014:023 1 12 01.00 06.13 06.13 00.80
015:076 1 03 01.00 06.13 06.13 01.32
016:029 1 08 01.00 06.13 06.13 00.97
016:080 1 20 01.00 06.13 06.13 00.59
018:108 1 03 01.00 06.13 06.13 01.32
019:073 1 07 01.00 06.13 06.13 01.02
020:071 1 19 01.00 06.13 06.13 00.61
020:073 1 07 01.00 06.13 06.13 01.02
020:076 1 08 01.00 06.13 06.13 00.97
020:101 1 06 01.00 06.13 06.13 01.08
020:131 1 10 01.00 06.13 06.13 00.87
021:008 1 06 01.00 06.13 06.13 01.08
021:034 2 05 02.00 06.13 12.26 01.51
021:099 1 02 01.00 06.13 06.13 01.42
021:102 1 06 01.00 06.13 06.13 01.08
023:011 1 03 01.00 06.13 06.13 01.32
023:103 1 07 01.00 06.13 06.13 01.02
025:015 1 07 01.00 06.13 06.13 01.02
025:016 1 06 01.00 06.13 06.13 01.08
025:069 1 06 01.00 06.13 06.13 01.08
025:076 1 04 01.00 06.13 06.13 01.23
029:058 1 13 01.00 06.13 06.13 00.76
031:009 1 06 01.00 06.13 06.13 01.08
032:014 1 09 01.00 06.13 06.13 00.92

033:065 1 06 01.00 06.13 06.13 01.08
035:035 1 09 01.00 06.13 06.13 00.92
039:072 1 07 01.00 06.13 06.13 01.02
039:073 1 13 01.00 06.13 06.13 00.76
039:074 1 13 01.00 06.13 06.13 00.76
040:076 1 07 01.00 06.13 06.13 01.02
041:028 1 09 01.00 06.13 06.13 00.92
043:071 1 09 01.00 06.13 06.13 00.92
043:074 1 05 01.00 06.13 06.13 01.15
046:014 1 04 01.00 06.13 06.13 01.23
047:015 1 32 01.00 06.13 06.13 00.42
047:019 1 11 01.00 06.13 06.13 00.83
048:005 1 14 01.00 06.13 06.13 00.73
050:034 1 04 01.00 06.13 06.13 01.23
057:012 1 18 01.00 06.13 06.13 00.63
058:017 1 07 01.00 06.13 06.13 01.02
058:022 1 29 01.00 06.13 06.13 00.45
059:017 1 05 01.00 06.13 06.13 01.15
064:009 1 20 01.00 06.13 06.13 00.59
064:010 1 08 01.00 06.13 06.13 00.97
065:011 1 20 01.00 06.13 06.13 00.59
072:023 1 11 01.00 06.13 06.13 00.83
098:006 1 10 01.00 06.13 06.13 00.87
098:008 1 12 01.00 06.13 06.13 00.80

/89

abidingplac

025:024 1 05 01.00 11.61 11.61 01.15
032:019 1 04 01.00 11.61 11.61 01.23

/2

abil

002:286 1 33 01.00 10.02 10.02 00.41
006:135 1 10 01.00 10.02 10.02 00.87
006:152 1 23 01.00 10.02 10.02 00.54
007:042 1 09 01.00 10.02 10.02 00.92
011:093 1 11 01.00 10.02 10.02 00.83
023:062 1 10 01.00 10.02 10.02 00.87

/6

abject

007:013 1 06 01.00 11.02 11.02 01.08

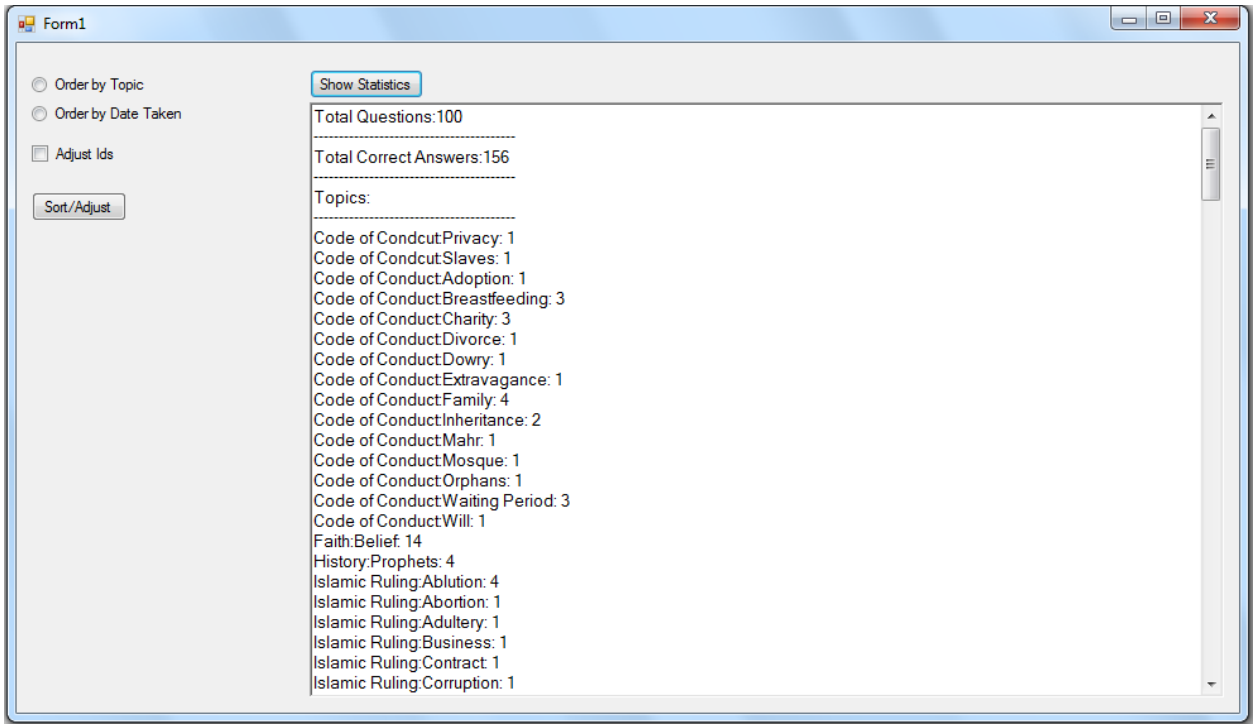
036:068 1 08 01.00 11.02 11.02 00.97
037:018 1 02 01.00 11.02 11.02 01.42
/3
ablut
005:006 1 37 01.00 12.61 12.61 00.38
/1
abod
002:036 1 11 01.00 06.38 06.38 00.83
002:094 1 08 01.00 06.38 06.38 00.97
003:151 2 11 02.00 06.38 12.76 01.21
003:162 1 08 01.00 06.38 06.38 00.97
003:197 1 05 01.00 06.38 06.38 01.15
004:097 1 16 01.00 06.38 06.38 00.68
004:121 1 04 01.00 06.38 06.38 01.23
005:072 1 19 01.00 06.38 06.38 00.61
006:032 1 10 01.00 06.38 06.38 00.87
006:127 1 04 01.00 06.38 06.38 01.23
006:128 1 25 01.00 06.38 06.38 00.51
006:135 1 10 01.00 06.38 06.38 00.87
007:024 1 05 01.00 06.38 06.38 01.15
007:078 1 05 01.00 06.38 06.38 01.15
007:091 1 05 01.00 06.38 06.38 01.15
007:145 1 16 01.00 06.38 06.38 00.68
007:169 1 20 01.00 06.38 06.38 00.59
008:016 1 15 01.00 06.38 06.38 00.70
009:072 1 20 01.00 06.38 06.38 00.59
009:073 1 10 01.00 06.38 06.38 00.87
009:095 1 10 01.00 06.38 06.38 00.87
010:008 1 03 01.00 06.38 06.38 01.32
010:025 1 06 01.00 06.38 06.38 01.08
010:093 1 16 01.00 06.38 06.38 00.68
011:065 1 06 01.00 06.38 06.38 01.08
011:067 1 06 01.00 06.38 06.38 01.08
011:094 1 12 01.00 06.38 06.38 00.80
012:021 1 17 01.00 06.38 06.38 00.65
012:023 1 19 01.00 06.38 06.38 00.61
012:109 1 11 01.00 06.38 06.38 00.83
013:018 1 14 01.00 06.38 06.38 00.73
013:022 1 15 01.00 06.38 06.38 00.70

013:023 1 11 01.00 06.38 06.38 00.83
013:024 1 05 01.00 06.38 06.38 01.15
013:025 1 14 01.00 06.38 06.38 00.73
013:031 1 27 01.00 06.38 06.38 00.48
013:042 1 09 01.00 06.38 06.38 00.92
014:028 1 09 01.00 06.38 06.38 00.92
014:045 1 07 01.00 06.38 06.38 01.02
016:030 2 13 02.00 06.38 12.76 01.13
016:041 1 10 01.00 06.38 06.38 00.87
017:097 1 19 01.00 06.38 06.38 00.61
024:057 1 07 01.00 06.38 06.38 01.02
025:066 1 06 01.00 06.38 06.38 01.08
025:076 1 04 01.00 06.38 06.38 01.23
028:037 1 09 01.00 06.38 06.38 00.92
028:058 1 08 01.00 06.38 06.38 00.97
028:077 1 16 01.00 06.38 06.38 00.68
028:081 1 08 01.00 06.38 06.38 00.97
028:083 1 12 01.00 06.38 06.38 00.80
029:025 1 11 01.00 06.38 06.38 00.83
029:037 1 07 01.00 06.38 06.38 01.02
029:058 1 13 01.00 06.38 06.38 00.76
029:064 1 07 01.00 06.38 06.38 01.02
029:068 1 08 01.00 06.38 06.38 00.97
032:020 1 11 01.00 06.38 06.38 00.83
032:026 1 07 01.00 06.38 06.38 01.02
033:029 1 09 01.00 06.38 06.38 00.92
034:015 1 13 01.00 06.38 06.38 00.76
038:046 1 08 01.00 06.38 06.38 00.97
039:032 1 08 01.00 06.38 06.38 00.97
039:060 1 08 01.00 06.38 06.38 00.97
039:072 1 07 01.00 06.38 06.38 01.02
040:039 1 08 01.00 06.38 06.38 00.97
040:052 1 07 01.00 06.38 06.38 01.02
040:076 1 07 01.00 06.38 06.38 01.02
041:024 1 06 01.00 06.38 06.38 01.08
045:034 1 08 01.00 06.38 06.38 00.97
047:012 1 15 01.00 06.38 06.38 00.70
057:015 1 09 01.00 06.38 06.38 00.92
059:009 1 15 01.00 06.38 06.38 00.70

066:009 1 10 01.00 06.38 06.38 00.87
079:039 1 03 01.00 06.38 06.38 01.32
079:041 1 03 01.00 06.38 06.38 01.32
101:009 1 02 01.00 06.38 06.38 01.42
/75
abomin
019:089 1 03 01.00 11.61 11.61 01.32
021:074 1 12 01.00 11.61 11.61 00.80
/2
abound
038:029 1 09 01.00 11.61 11.61 00.92
050:009 1 09 01.00 11.61 11.61 00.92
/2
abroad
004:083 1 14 01.00 11.02 11.02 00.73
045:004 1 06 01.00 11.02 11.02 01.08
062:010 1 10 01.00 11.02 11.02 00.87
/3
abrog
002:106 1 06 01.00 12.61 12.61 01.08
/1
absenc
002:092 1 07 01.00 12.61 12.61 01.02
/1
absent
007:007 1 03 01.00 11.02 11.02 01.32
034:003 1 14 01.00 11.02 11.02 00.73
082:016 1 02 01.00 11.02 11.02 01.42
/3
absente
027:020 1 04 01.00 12.61 12.61 01.23
/1
absolut
030:007 1 05 01.00 11.02 11.02 01.15
033:038 1 08 01.00 11.02 11.02 00.97
052:037 1 05 01.00 11.02 11.02 01.15
/3

Appendix 5 – PARMS Supporting Tools

Question Helper Tool



Appendix 6 – PARMS Testing Evidence

Test Question

What is the law of retribution and Qisas?

Punctuation removal

What is the law of retribution and Qisas

Stopwords identification – identified stopwords are crossed out and will be ignored in following steps.

~~*What*~~ ~~*is*~~ ~~*the*~~ *law* ~~*of*~~ *retribution* ~~*and*~~ *Qisas*

Question Tokens

law retribution Qisas

Question Enhancement – synonyms and QurTems related to each token in the question. Duplicates and stopwords are crossed out and will be ignored in following steps:

Token	Synonyms	QurTerms
Law	jurisprudence	
Retribution	*Retribution*, requital	
Qisas		*Blood money*, ~~*Blood*~~ *ransom*, ~~*Blood*~~ *compensation*, *Innocent Killing payment*, *Retaliation*, *Equitable* ~~*retaliation,*~~ ~~*Retaliation*~~ ~~*of*~~ ~~*blood,*~~ *Law* ~~*of*~~ *equality* ~~*in*~~ *punishment*

Enhanced Tokens (Question + synonyms + Qur Terms)

law jurisprudence *retribution* requital q*isas blood money ransom compensation innocent killing payment retaliation equitable law equality punishment*

Tokens Stemming:

Token	Stemmed Token
law:	law
jurisprudence:	jurisprud
retribution:	retribut
requital:	requit
qisas:	qisa
blood:	blood
money:	money
ransom:	ransom
compensation:	compens
innocent:	innoc
killing:	kill
payment:	payment
retaliation:	retali

```
equitable:        equit
law:              law
equality:         equal
punishment:       punish
```

Final Test Question Terms: Following table provides the list of tokens along with their corresponding term and term frequency:

Token	Term	Frequency
law	law	2
Jurisprudence	jurisprud	1
Retribution	retribut	1
Requital	requit	1
Qisas	qisa	1
blood	blood	1
money	money	1
ransom	ransom	1
compensation	compens	1
Innocent	innoc	1
killing	kill	1
payment	payment	1
Retaliation	retali	1
Equitable	equit	1
equality	equal	1
punishment	punish	1

Test Answer: (Corpus Ref: Shakir 002:178)

O you who believe! retaliation is prescribed for you in the matter of the slain, the free for the free, and the slave for the slave, and the female for the female, but if any remission is made to any one by his (aggrieved) brother, then prosecution (for the bloodwit) should be made according to usage, and payment should be made to him in a good manner; this is an alleviation from your Lord and a mercy; so whoever exceeds the limit after this he shall have a painful chastisement.

Punctuation Removal:

O you who believe retaliation is prescribed for you in the matter of the slain the free for the free and the slave for the slave and the female for the female but if any remission is made to any one by his aggrieved brother then prosecution for the bloodwit should be made according to usage and payment should be made to him in a good manner this is an alleviation from your Lord and a mercy so whoever exceeds the limit after this he shall have a painful chastisement

Stopwords identification – identified stop words are crossed out and will be ignored in following steps:

~~*O you who believe*~~ *retaliation* ~~*is*~~ *prescribed* ~~*for you in the*~~ *matter* ~~*of the*~~ *slain* ~~*the*~~ *free* ~~*for the*~~ *free* ~~*and the*~~ *slave* ~~*for the*~~ *slave* ~~*and the*~~ *female* ~~*for the*~~ *female* ~~*but if any*~~ *remission* ~~*is*~~ *made* ~~*to any one by his*~~ *aggrieved brother* ~~*then*~~ *prosecution* ~~*for the*~~ *bloodwit* ~~*should be*~~ *made* ~~*according to*~~ *usage* ~~*and*~~ *payment* ~~*should be made to him in a*~~ *good manner* ~~*this is an*~~ *alleviation* ~~*from your*~~ *Lord* ~~*and a*~~ *mercy* ~~*so whoever*~~ *exceeds* ~~*the*~~ *limit* ~~*after this he shall have a*~~ *painful chastisement*

Answer Tokens

retaliation prescribed matter slain free free slave slave female female remission made aggrieved brother prosecution bloodwit made usage payment made good manner alleviation lord mercy exceeds limit painful chastisement

Answer Enhancement - pronouns added from Qurana, stopwords are crossed out and will be ignored in following steps**:**

null, ~~*those who believe*~~*,* ~~*the*~~ *Murderer,* ~~*the*~~ *murdered, guardians* ~~*of the*~~ *Killed,* ~~*those who believe*~~*, transgressor*

Answer Tokens (answer + pronouns)

retaliation prescribed matter slain free free slave slave female female remission made aggrieved brother prosecution bloodwit made usage payment made good manner alleviation lord mercy exceeds limit painful chastisement null Murderer murdered guardians Killed transgressor

Stemmed Tokens:

Token	**Stemmed Token**
retaliation:	*retali*
prescribed:	*prescrib*
matter:	*matter*
slain:	*slain*
free:	*free*
free:	*free*
slave:	*slave*
slave:	*slave*
female:	*femal*
female:	*femal*
remission:	*remiss*
made:	*made*
aggrieved:	*aggriev*
brother:	*brother*
prosecution:	*prosecut*
bloodwit:	*bloodwit*
made:	*made*
usage:	*usag*
payment:	*payment*
made:	*made*
good:	*good*
manner:	*manner*
alleviation:	*allevi*
lord:	*lord*
mercy:	*merci*
exceeds:	*exceed*
limit:	*limit*
painful:	*pain*
chastisement:	*chastis*
null:	*null*
murderer:	*murder*
murdered:	*murder*
guardians:	*guardian*
killed:	*kill*
transgressor:	*transgressor*

Final Test Answer Terms: Following table provides the list of tokens along with their corresponding term and term frequency:

Token	**Term**	**Frequency**
retaliation	*retali*	1
prescribed	*prescrib*	1
matter	*matter*	1
slain	*slain*	1

Token	Term	Frequency
free	*free*	2
slave	*slave*	2
female	*femal*	2
remission	*remiss*	1
made	*made*	3
aggrieved	*aggriev*	1
brother	*brother*	1
prosecution	*prosecut*	1
bloodwit	*bloodwit*	1
usage	*usag*	1
payment	*payment*	1
good	*good*	1
manner	*manner*	1
alleviation	*allevi*	1
lord	*lord*	1
mercy	mercy	1
exceeds	*exceed*	1
limit	*Limit*	1
painful	pain	1
chastisement	*chastis*	1
null	null	1
murderer *murdered*	*murder*	2
guardians	*guardian*	1
killed	kill	1
transgressor	*transgressor*	1
Total Terms		**35**

Terms matching with Question terms & Weights:

Total Documents: 6236

Term	Number of documents terms exists	idf
retali	5	$Log_2(6236/5) = 10.28447711$
payment	4	$Log_2(6236/4) = 10.60640521$
kill	49	$Log_2(6236/49) = 6.99169536$
Document Weight:		**27.88257768**

Answer: (Corpus Ref: Shakir 005:095)

O you who believe! do not kill game while you are on pilgrimage, and whoever among you shall kill it intentionally, the compensation (of it) is the like of what he killed, from the cattle, as two just persons among you shall judge, as an offering to be brought to the Kaaba or the expiation (of it) is the feeding of the poor or the equivalent of it in fasting, that he may taste the unwholesome result of his deed; Allah has pardoned what is gone by; and whoever returns (to it), Allah will inflict retribution on him; and Allah is Mighty, Lord of Retribution.

Punctuation Removal:

O you who believe do not kill game while you are on pilgrimage and whoever among you shall kill it intentionally the compensation of it is the like of what he killed from the cattle as two just persons among you shall judge as an offering to be brought to the Kaaba or the expiation of it is the feeding of the poor or the equivalent of it in fasting that he may taste the unwholesome result of his deed Allah has pardoned what is gone by and whoever returns to it Allah will inflict retribution on him and Allah is Mighty Lord of Retribution

Stopwords identification – identified stop words are crossed out and will be ignored in following steps

~~O you who believe do not~~ kill game ~~while you are on~~ pilgrimage ~~and whoever among you shall~~ kill ~~it~~ intentionally ~~the~~ compensation ~~of it is the like of what he~~ killed ~~from the~~ cattle ~~as two just~~ persons ~~among you shall~~ judge ~~as an~~ offering ~~to be~~ brought ~~to the~~ Kaaba ~~or the~~ expiation ~~of it is the~~ feeding ~~of the~~ poor ~~or the~~ equivalent ~~of it in~~ fasting ~~that he may~~ taste ~~the~~ unwholesome result ~~of his~~ deed ~~Allah~~ ~~has~~ pardoned ~~what is gone by and whoever~~ returns ~~to it Allah will~~ inflict retribution ~~on him and Allah is~~ Mighty Lord ~~of~~ Retribution

Answer Tokens:

kill game pilgrimage kill intentionally compensation killed cattle persons judge offering brought kaaba expiation feeding poor equivalent fasting taste unwholesome result deed pardoned returns inflict retribution mighty lord retribution

Answer Enhancement - pronouns added from Qurana, stopwords are crossed out and will be ignored in following steps

~~those, who, believe, those, who, believe, those, who, believe~~, hunted, animals, ~~those, who, believe~~, penalty, ~~of~~, hunting, ~~while~~, wearing, Ihram, ~~those, who, believe, who~~, hunts, ~~in, the~~, state, ~~of~~, Ihram, ~~who~~, hunts, ~~in, the~~, state, ~~of~~, Ihram

Answer Tokens (answer + pronouns)

kill game pilgrimage kill intentionally compensation killed cattle persons judge offering brought kaaba expiation feeding poor equivalent fasting taste unwholesome result deed pardoned returns inflict retribution mighty lord retribution hunted animals penalty hunting wearing Ihram hunts state Ihram hunts state Ihram

Stemmed Tokens:

Token	Stemmed Token
kill:	kill
game:	game
pilgrimage:	pilgrimag
kill:	kill
intentionally:	intent
compensation:	compens
killed:	kill
cattle:	cattl
persons:	person
judge:	judg
offering:	offer
brought:	brought
kaaba:	kaaba
expiation:	expiat
feeding:	feed
poor:	poor
equivalent:	equival
fasting:	fast
taste:	tast
unwholesome:	unwholesom
result:	result
deed:	deed
pardoned:	pardon
returns:	return
inflict:	inflict
retribution:	retribut
mighty:	mighti
lord:	lord
retribution:	retribut
hunted:	hunt
animals:	anim
penalty:	penalti
hunting:	hunt
wearing:	wear
ihram:	ihram
hunts:	hunt
state:	state
ihram:	ihram
hunts:	hunt
state:	state
ihram:	ihram

Final Test Answer Terms: Following table provides the list of tokens along with their corresponding term and term frequency:

Token	Term	Frequency
Kill, killed	kill	3
game	game	1

pilgrimage	pilgrimag	1
intentionally	intent	1
compensation	compens	1
cattle	cattl	1
persons	person	1
judge	judg	1
offering	offer	1
brought	brought	1
kaaba	kaaba	1
expiation	expiat	1
feeding	feed	1
poor	poor	1
equivalent	equival	1
fasting	fast	1
taste	tast	1
unwholesome	unwholesom	1
result	result	1
deed	deed	1
pardoned	pardon	1
returns	return	1
inflict	inflict	1
retribution	retribut	2
mighty	mighti	1
lord	lord	1
Hunted, hunting, hunts	hunt	4
animals	anim	1
penalty	penalti	1
wearing	wear	1
ihram	ihram	3
state	state	2

Terms matching with Question & Weights:

Total Documents: 6236

Term	**Number of documents terms exists**	**idf**
retribut	22	$Log_2(6236/22) = 8.14697359$
compens	5	$Log_2(6236/5) = 10.28447711$
kill	49	$Log_2(6236/49) = 6.99169536$
Document Weight:		**25.42314606**

Bibliography

Abbas, N. (2009). "The Qurany Tool - Keyword Search." Retrieved 20 Feb, 2013, from http://quranytopics.appspot.com/.
Abbott, J. and T. Ryan (2000). The Unfinished Revolution. Learning, Human Behaviour, Community and Political Paradox.
Abdelsalam, A. M., B. S. Farouk, R. M. Samir, M. I. Ragab and H. A. EL-Nasser (2013). Quran Expert - Al Bayan. BSc Degree Graduation Project, Alexandria University.
Abdul-Baquee. (2013, 20 July 2012). "TextMiningTheQuran.Com." 2011. Retrieved 20 Feb, 2013, from http://www.textminingthequran.com/apps/conceptlist.php.
Aisa Mustapha, e. a. (2012). "The Qur'an Corpus for Juz' Amma ".
Al-Munajjid, S. M. S. (2013). "IslamQA.com." 1987. Retrieved 20 Feb, 2013, from http://islamqa.com/en/tree.
Ali, A. (2003). "Study the Quran." from http://studythequran.com/answers_from_the_holyquran.php.
Ali, A. (2013). "StudytheQuran.com." 2003. Retrieved 20 Feb, 2012, from http://studythequran.com/answers_from_the_holyquran.php.
Ali, A., M. A. Brakhw, M. Z. F. B. Nordin and S. F. Shaik (2012). "Some Linguistic Difficulties in Translating the Holy Quran from Arabic into English." International Journal of Social Science and Humanity, **2**(6): 588 - 590.
Amjad, M. (2001). "Understanding Islam." 1997. from http://forums.understanding-islam.com/.
Ammon, U. (2001). The dominance of English as a language of science: Effects on other languages and language communities, De Gruyter Mouton.
Anon. (2002). "Islamic Knowledge - Authenticity of Islamic Information." from http://islamic-knowledge.com/Books_Articles/Authenticity_of_Islamic_Information.htm.
Attia, M. A. (2008). Handling Arabic Morphological and Syntactic Ambiguity within the LFG Framework with a View to Machine Translation, the University of Manchester.
Atwell, E., C. Brierley, K. Dukes, M. Sawalha and A.-B. Sharaf (2011). An Artificial Intelligence approach to Arabic and Islamic content on the internet. Proc NITS'2011 National Information Technology Symposium, King Saud University, Saudi Arabia. Data protection statements.
Atwell, E., C. Brierly and M. Sawalha (2012). Introduction to Language Resources and Evaluation for Religious Texts. Language Resources and Evaluation for Religious Texts, Istanbul, Turkey.
Atwell, E. e. a. (2010). "Understanding the Quran: A New Grand Challenge for Computer Science and Artificial Intelligence." ACM/BCS Visions of Computer Science.
Aznan Zuhid Saidin, e. a. (2005). The Code of Ethics for Software Engineers, Public Interest and the Islamic Perspective. Prosiding Seminar Kebangsaan E-Komuniti, Putrajaya.
Bakalla, M. (2002). Arabic culture through its language and literature, Kegan Paul International.
Baker, M. (2008). Routledge encyclopedia of translation studies, Routledge.
Beaulieu, M., M. Gatford, X. Huang, S. Robertson, S. Walker and P. Williams (1997). "Okapi at TREC-5." NIST SPECIAL PUBLICATION SP: 143-166.
Bewley, A. A. (1999). The Nobel Quran: A New Rendering of its Meaning in English. Norwich, UK, Bookwork
Brewster, C. and K. O'Hara (2004). "Knowledge representation with ontologies: the present and future." Intelligent Systems, IEEE **19**(1): 72-81.
Caplan, S. (2013) "Tongue-tied? Perspectives on English as the international language of science." Ocaam's Corner.
Choueiri, Y. M. (2010). Islamic Fundamentalism 3rd Edition: The Story of Islamist Movements, Continuum.
Christmann, A. (2002). "The Noble Quran: A New Rendering of its Meaning in English." Journal of Semitic Studies **47**(2): 372-374.
Correa, S., D. Buscaldi and P. Rosso (2010). NLEL at RespubliQA 2010. CLEF (Notebook Papers/LABs/Workshops).
Cudden, J. A., Ed. (1992). The Dictionary of Literary Terms and Literary Theory

London: .
D. M. Aliod, J. L. V. (2010). Question Answering. Machine Learning & Pattern Recognition Series. F. J. D. N. Indurkhya. New York, USA, Chapman & Hall/CRC**:** 675.
Di Cai and A. J. Lee McCluskey (2011). Rummage System Report. Huddersfield, UK, The University of Huddersfield**:** 18.
Dukes, K. (2011). "The Quranic Arabic Corpus." 2009. 2013, from http://corpus.quran.com/.
Dukes, K. and E. Atwell (2011). "LAMP: A Multimodal Web Platform for Collaborative Linguistic Analysis."
Eric Atwell, e. a. (2013). "Natural Language Processing Working Together with Arabic and Islamic Studies." Retrieved February 2013, 2013, from http://www.engineering.leeds.ac.uk/computing/events/upcoming.shtml.
Farghaly, A. and K. Shaalan (2009). "Arabic natural language processing: Challenges and solutions." ACM Transactions on Asian Language Information Processing (TALIP) **8**(4): 14.
Fellbaum, C. (1998). WordNet, Wiley Online Library.
Ferguson, C. (2006). Epilogue: Diglossia revisited. Understanding Arabic: Essays in Contemporary Arabic Linguistics in Honor of El-Said M. Badawi. A. E. El-Said M. Badawi, The American University in Cairo.
Glöckner, I. and B. Pelzer (2010). The LogAnswer Project at ResPubliQA 2010. CLEF (notebook papers/labs/workshops).
Hammo, B., H. Abu-Salem and S. Lytinen (2002). QARAB: A question answering system to support the Arabic language. Proceedings of the ACL-02 workshop on Computational approaches to semitic languages, Association for Computational Linguistics.
Hartman, E. M. (2008). Socratic questions and Aristotelian answers: A virtue-based approach to business ethics. Leadership and Business Ethics, Springer**:** 81-101.
Hersch, R. (1998). "FAQ about FAQ's." Retrieved 11th June, 2013, from http://www.faqs.org/faqs/faqs/about-faqs/.
Imam. (2013). "Islam City." 1995. from http://www.islamicity.com.
Ismail, H. (2009). Search Engine System for the Holy Quran. Thesis (Masters) University Utara, Malaysia.
ISOC. (1980). "The Islamic Society of Huddersfield University." from www.hud-isoc.org.
Jilani, A. (2012). PARallel Multi-Stream QA System. Knowledge Engineering Intelligent Interfaces Research Group Presentation. A. Jilani. Huddersfield, University of Huddersfield.
Jilani, A. (2013). PARallel Multi-Stream Methodology. Knowledge Engineering Intelligent Interfaces Research Group Presentation. A. Jilani. Huddersfield, University of Huddersfield.
Jilani, A. (2013). Quran Ontologies and Keywords for Question Answering. Proceedings of WACL#'2 Second Workshop on Arabic Corpus Linguistics, Lancaster University, Lancaster, UK.
Jilani, A. (2013). Quran Ontologies and Keywords for Question Answering. Second Workshop on Arabic Corpus Linguistics. A. H. Eric Atwell. Lancaster University, UK.
Jones, H. (2011) "2.2 Billion: World's Muslim Population Doubles."
Jones, K. S. (1972). "A statistical interpretation of term specificity and its application in retrieval." Journal of documentation **28**(1): 11-21.
Jones, K. S. r. (2004). "IDF term weighting and IR research lessons." Journal of Documentation **60**(5): 521-523.
Kaplan, R. B. (2001). "English–the accidental language of science." The Dominance of English as a Language of Science: Effects on Other Languages and Language Communities. Berlin: 3-26.
Keller, S. N. (2013). "Qibla Answers." from http://spa.qibla.com/.
Kettani, H. (2010). 2010 World Muslim Population. proceedings of the 8th Hawaii Internafional Conference on Arts and Humanifies.
Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. Proceedings of the 5th annual international conference on Systems documentation. Toronto, Ontario, Canada, ACM**:** 24-26.
Lin, J. and B. Katz (2006). "Building a reusable test collection for question answering." Journal of the American Society for Information Science and Technology **57**(7): 851-861.
Manning, C. D., P. Raghavan and H. Schütze (2008). Introduction to information retrieval, Cambridge University Press Cambridge.

Marican, M. (2013). "Hajj Questions and Answers." from http://www.hajtips.com/faqs.htm.
Meenachi, N. M. and M. Sai Baba (2012). "A Survey on usage of Ontology in Different Domains." International Journal of Applied Information Systems **9**(2): 46-55.
Miller, G. A. (1995). "WordNet: a lexical database for English." Communications of the ACM **38**(11): 39-41.
Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross and K. J. Miller (1990). "Introduction to wordnet: An on-line lexical database*." International journal of lexicography **3**(4): 235-244.
Mohammed, K. (2005). "Assessing English translations of the Qur'an." Middle East Quarterly.
Mohd Yunus, M. A. (2008). Short Query Translation - A dictionary Based Approach to Cross Language Information Retrieval. Masters, Universiti Teknologi MARA.
Mohideen, H. (2007). "Islamic Terms in Contemporary English." English Today **23**(2).
Monz, C. (2003). From document retrieval to question answering, Institute for Logic, Language and Computation.
Morato, J., M. A. Marzal, J. Lloréns and J. Moreiro (2004). WordNet applications. Proceedings of the 2nd Global Wordnet Conference.
Mufti. (2013). "Islam Web (English)." 1998. from http://www.islamweb.net/emainpage/index.php?page=fatwa.
Muhammad, A. B. (2013). Annotation of Conceptual Co-reference and Text Mining the Qur'an. Doctor of Philosophy, The University of Leeds.
Mustapha, A. (2009). "Dialogue-based visualization for quranic text." European Journal of Scientific Research **37**(1): 36-40.
Mutahhari, M. (1983) "Understanding the Uniqueness of the Qur'an " **1-3**.
New, B. (2008). Question Answering at TREC Pre-Internship Report. Ling 590, University of Washington.
NIST. (1992). "The Text REtrieval Conference (TREC)." from http://trec.nist.gov/overview.html.
Noordin, M. F. and R. Othman (2006). An Information Retrieval System for Quranic Texts: A Proposed System Design. Information and Communication Technologies, 2006. ICTTA'06. 2nd, IEEE.
Oktar, A. (2013). "Harun Yahya works." 1994. from http://www.harunyahya.com/list/type/6/name/Articles/category/78/orderBy/1/limit/100.
Padesky, C. A. (1993). Socratic questioning: Changing minds or guiding discovery. A keynote address delivered at the European Congress of Behavioural and Cognitive Therapies, London.
Pakray, P., P. Bhaskar, S. Pal, D. Das, S. Bandyopadhyay and A. F. Gelbukh (2010). JU CSE TE: System Description QA@ CLEF 2010-ResPubliQA. CLEF (Notebook Papers/LABs/Workshops).
Pearsall, J. and P. Hanks (1998). The new Oxford dictionary of English, Clarendon Press.
Pedersen, T., S. Patwardhan and J. Michelizzi (2004). WordNet:: Similarity: measuring the relatedness of concepts. Demonstration Papers at HLT-NAACL 2004, Association for Computational Linguistics.
Peñas, A., P. Forner, R. Sutcliffe, Á. Rodrigo, C. Forăscu, I. Alegria, D. Giampiccolo, N. Moreau and P. Osenova (2010). Overview of ResPubliQA 2009: Question answering evaluation over European legislation. Multilingual Information Access Evaluation I. Text Retrieval Experiments, Springer**:** 174-196.
Pérez-Iglesias, J., G. Garrido, Á. Rodrigo, L. Araujo and A. Peñas (2010). Information retrieval baselines for the ResPubliQA task. Multilingual Information Access Evaluation I. Text Retrieval Experiments, Springer**:** 253-256.
PEWResearch (2011) "The Future of the Global Muslim Population."
Plachouras, V., B. He and I. Ounis (2004). University of Glasgow at TREC 2004: Experiments in Web, Robust, and Terabyte Tracks with Terrier. TREC.
Rahim, A. (2007). "Studying Quran." 2003. from http://www.studying-islam.org/query.aspx.
Resnik, P., M. B. Olsen and M. Diab (1999). "The Bible as a parallel corpus: Annotating the 'Book of 2000 Tongues'." Computers and the Humanities **33**(1): 129-153.

Rhodora (2012). "The Structure of Entropy - Qur'an Anthologies: Illuminating Metaphors." http://structureofentropy.wordpress.com/2012/07/15/quran-anthologies-illuminating-metaphors/.
Robertson, S. E. and S. Walker (1999). Okapi/Keenbow at TREC-8. TREC.
Robertson, S. E., S. Walker, M. Beaulieu and P. Willett (1999). "Okapi at TREC-7: automatic ad hoc, filtering, VLC and interactive track." Nist Special Publication SP: 253-264.
Robertson, S. E., S. Walker, M. Hancock-Beaulieu, A. Gull and M. Lau (1993). Okapi at TREC. Text retrieval conference.
Rodrigo, Á., J. Perez-Iglesias, A. Peñas, G. Garrido and L. Araujo (2010). A Question Answering System based on Information Retrieval and Validation. CLEF (Notebook Papers/LABs/Workshops).
Rusbergen, C. J. v. (1979). Information Retrieval, Butterworth.
Saad, S., N. Salim, H. Zainal and Z. Muda (2011). A process for building domain ontology: An experience in developing Solat ontology. Electrical Engineering and Informatics (ICEEI), 2011 International Conference on, IEEE.
Sabnani, H. and P. Majumder (2010). Question Answering System: Retrieving Relevant Passages. CLEF (Notebook Papers/LABs/Workshops), Citeseer.
Salhey. (2013, 20 Feb 2013). "Quran.al-Islam." Mauqa.al-Islam, from http://quran.al-islam.com/Loader.aspx?pageid=706.
Salton, G. and C. Buckley (1988). "Term-weighting approaches in automatic text retrieval." Information processing & management **24**(5): 513-523.
Salton, G., E. A. Fox and H. Wu (1983). "Extended Boolean information retrieval." Communications of the ACM **26**(11): 1022-1036.
Salton, G. and M. J. McGill (1983). "Introduction to modern information retrieval."
Sanderson, M. and A. Alberair (2001). "Keep It Simple Sheffield–a KISS approach to the Arabic track." NIST Special Publication 500-250: 642-643.
Sawalha, M. and E. Atwell (2010). Fine-Grain Morphological Analyzer and Part-of-Speech Tagger for Arabic Text. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA).
Schlaefer, N. (2007). "OpenEphyra." from http://www.ephyra.info/.
Schlenker, P. (2006). Introduction to the Study of Language - Sentence Structure I: Syntactic Trees**:** 8.
Scholars. (2013). "On Islam." 2000 - FB: 2010. from http://onislam.net/english/ask-the-scholar.html.
Shah, S. N. (2010). "Alim.org." 1986. 2013, from http://www.alim.org/faq.
Sharaf, A.-B. M. and E. Atwell (2012). "QurAna: Corpus of the Quran annotated with Pronominal Anaphora." LREC 2012.
Sharaf, A.-B. M. and E. Atwell (2012). "QurSim: A corpus for evaluation of relatedness in short texts." LREC 2012.
Sheikh, F. T. (2006). "Qur'anic Imagery of Doomsday and Resurrection." Renaissance **Vol. 16** (Issue 10).
Sherif, M. A. and A.-C. N. Ngomo (2009). "Semantic Quran." IOS Press.
Singhal, A. (2001). "Modern information retrieval: A brief overview." IEEE Data Eng. Bull. **24**(4): 35-43.
Smith, C. S. (2009). "Parallel English Quran translation." Retrieved 20 Feb 2013, from http://clay.smith.name/Parallel_Quran.htm.
Soltan. (2012). "FAQ about Islam." from http://sultan.org/articles/ISLAMFAQ.html.
Stevenson, A. (2010). Oxford dictionary of English, Oxford University Press.
Sultan, A. M. (2011). Evaluation of Quranic text retrieval system based on manually indexed topics. Semantic Technology and Information Retrieval (STAIR), 2011 International Conference on, IEEE.
Sultan, A. M., A. Azman, R. A. Kadir and M. T. Abdullah (2011). Evaluation of Quranic text retrieval system based on manually indexed topics. Semantic Technology and Information Retrieval (STAIR), 2011 International Conference on, IEEE.
Svenonius, E. (1972). "An experiment in index term frequency." Journal of the American Society for Information Science **23**(2): 109-121.
Taha Zerrouki, e. a. (2012). Reusability of Quranic Document using XML. Language Resources and Evaluation for Religious Texts, Istanbul, Leeds University.

Twilightdreamlover (2008). "It's the HUMAN thing to do!!" AskVille by Amazon http://askville.amazon.com/humans-compelling-questions/AnswerViewer.do?requestId=11793298.
Tzortzis, H. A. (2008) "Qur'ans Unique Literary Form."
Versteegh, K. and C. Versteegh (1997). The Arabic Language, Columbia University Press.
Voorhees, E. M. and D. M. Tice (2000). Building a question answering test collection. Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, ACM.
Whissell, C. (2004). "Emotion and Style in an English Translation of the Quran 1,2." Psychological reports **94**(2): 523-544.
Wu, H. C., R. W. P. Luk, K. F. Wong and K. L. Kwok (2008). "Interpreting tf-idf term weights as making relevance decisions." ACM Transactions on Information Systems (TOIS) **26**(3): 13.
Xu, F. and D. Klakow (2010). Paragraph Acquisition and Selection for List Question Using Amazon's Mechanical Turk. LREC.
Yang, Y.-T. C., T. J. Newby and R. L. Bill (2005). "Using Socratic questioning to promote critical thinking skills through asynchronous discussion forums in distance learning environments." The American Journal of Distance Education **19**(3): 163-181.
Yunus, M., R. Zainuddin and N. Abdullah (2010). Semantic query for Quran documents results. Open Systems (ICOS), 2010 IEEE Conference on, IEEE.
Yunus, M., R. Zainuddin and N. Abdullah (2010). Semantic query with stemmer for Quran documents results. Open Systems (ICOS), 2010 IEEE Conference on, IEEE.
Yunus, M., R. Zainuddin and N. Abdullah (2011). Semantic speech query via stemmer for Quran documents results. Electronic Devices, Systems and Applications (ICEDSA), 2011 International Conference on, IEEE.
Yusof, R., R. Zainuddin, M. Baba and Z. Yusoff (2009). "Visualization systems supporting the reading of arabic document for non arabic speakers." Information Technology Journal **8**(1): 16-27.
Zaidi S., e. a. (2012). Hybrid Approach for Extracting Collocations from Arabic Quran Texts. Proceedings of Workshop on Language Resources and Evaluation for Religious Texts.