



Sudan University of Science and Technology

College of Post Graduate Studies

An automatic question answering system for the Arabic Quran

نظام الرد الآلي على السؤال للقرآن العربي

A thesis submitted in fulfillment of the requirements of the award of the degree of

Doctor of Philosophy (Computer Science)

by: Mohamed Adany Hamdelsayed Adany

Supervisor: Eric S Atwell, Associate Professor

August 2017



Approval Page

(To be completed after the college council approval)

Name of Candidate:

Thesis title: AN AUTOMATIC QUESTION ANSWERING
SYSTEM FOR THE ARABIC QURAN.

Degree Examined for: PhD

Approved by:

1. External Examiner

Name: Iman Abdel Maady Abdelrahman

Signature: [Signature] Date: 7/8/2017

2. Internal Examiner

Name: Mohamed Elhafiz Mustafa Mada

Signature: [Signature] Date: 7/8/2017

3. Supervisor

Name: ERIK ATVEU

Signature: [Signature] Date: 7/8/2017

Declaration of the Status of Thesis (by student)

I, the signing here-under, declare that the contents of this dissertation represent my own work, and that the dissertation has not previously been submitted for academic examination towards any qualification.

Furthermore, it represents my own opinions, which is an original intellectual work.

Candidate's name: Mohamed Adany Hamdelsayed Adany

Candidate's signature:

Date:

Declaration of the Status of Thesis (by Supervisor)

I, the signing here-under, declare that I am the supervisor of the sole author of the Ph.D. dissertation entitled: An automatic question answering system for the Arabic Quran

Supervisor's name: Eric S Atwell , Associate Professor

Supervisor's signature:

Date

Assigning the copy-right to CGS

I, the signing here-under, declare that I am the sole author of the Ph.D. dissertation entitled: An automatic question answering system for the Arabic Quran. This is an original intellectual work. Willingly, I assign the copy-right of this work to the College of Graduate Studies (CGS), Sudan University of Science and Technology (SUST). Accordingly, SUST has all the rights to publish this work for scientific purposes.

Candidate's name: Mohamed Adany Hamdelsayed Adany

Candidate's signature:

Date:



Sudan University of Science and Technology
College of Post Graduate Studies



An automatic question answering system for the Arabic Quran

نظام الرد الآلي على السؤال للقرآن العربي

A thesis submitted in fulfillment of the requirements of the award of the degree of

Doctor of Philosophy (Computer Science)

by: Mohamed Adany Hamdelsayed Adany

Supervisor: Eric S Atwell, Associate Professor

August 2017

Introductive

قَالَ تَعَالَى:

﴿قُلْ لَوْ كَانَ الْبَحْرُ مَدَادًا لَكَلِمَتِ رَبِّي لَنَفِدَ الْبَحْرُ قَبْلَ أَنْ تَنفَدَ كَلِمَتُ رَبِّي وَلَوْ جِئْنَا بِمِثْلِهِ مَدَدًا﴾

الكهف: ١٠٩

"If the ocean were ink (wherewith to write out) the words of my Lord, sooner would the ocean be exhausted than would the words of my Lord, even if we added another ocean like it, for its aid."¹

The cave 109

¹ The Meanings Of The Holy Qur'an by Abdullah Yusuf ali

Dedication

I dedicate this to:

*My God, who delighted us this work Grace
of God and his strength.*

The spirit of my father and my brothers.

My mother.

My wife and children

*Prof Eric S Atwell for his patience and
his support to finish this thesis.*

*All my colleagues and my students who help
me to finish this modest work.*

*Mr. Eltayeb Elthamy Mohammed who
helped me to review some errors*

Acknowledgement

Firstly, I would like to express my honest thankfulness to my advisor Prof. Eric S Atwell for his support of my Ph.D study and related research, for his patience. His guidance helped me all the time to go the right way.

Abstract

This thesis investigates Question answering system in general and then applying it in a specific domain: Arabic language and Holy Quran. A corpus of questions and answers from Albagarah and Alfatihah chapters was built, using three types of files: excel worksheet, texts and database table. This corpus was collected from many resources and validated by Islamic scholars from Gabrah college. The thesis contains six chapters: Chapter one defines the problem statement, objectives and motivation, and research methodology. Chapter two gives a general introduction and history to question answering systems, natural language processing, and corpus. Chapter three reviews related work in Arabic language processing, Holy Quran, Corpus, question answering systems. Chapter four covers special characteristics and challenges of processing Arabic language in the Holy Quran. Chapter five contains methodology, methods, and experiments. Six sets of experiments were done. The first one uses baseline NLP question answering system removing stopwords, diacritics, and special symbols. The second uses Lucene indexing and تفاعيل pattern. The third uses indexing and تفاعيل، فاعيل، فياعيل pattern. The fourth uses the dynamic corpus built from real user questions. The fifth uses Exaggeration formulas pattern. The sixth uses singular, dual and plural pattern. Finally there are results and discussion. The experiments showed that removing stop words and diacritics enhanced the search results; also the new patterns added more value to the question answering system and enhanced recall and precision.

المستخلص

هذه الأطروحة تستكشف أنظمة الرد الآلي بصورة عامة ومن ثم تطبيقها في مجال خاص وهو اللغة العربية والقرآن الكريم تم بناء ذخيرة لغوية لبعض الأسئلة والأجوبة المقابلة لها من سورتي البقرة والفاحة؛ تم استخدام ثلاثة أنواع من الملفات ملف نصي، ورقة عمل (اكسيل)، جدول قاعدة بيانات. تم تجميع هذه الذخيرة من مصادر عديدة وتم تحكيم إجابات الأسئلة بواسطة علماء من كلية جيرة العلمية. يتكون هذا البحث من ستة أبواب : الفصل الأول يحتوي على

المشكلة والأهداف والدوافع، ومنهجية البحث. الفصل الثاني يعطي مقدمة عامة ولمحة تاريخية لأنظمة الرد الآلي ، ومعالجة اللغة الطبيعية، ومنهجية البحث. الفصل الثالث يستعرض الأعمال ذات الصلة في معالجة اللغة العربية، القرآن الكريم، الذخائر اللغوية ، وأنظمة الرد الآلي. ويغطي الفصل الرابع بعض الخصائص والتحديات الخاصة باللغة العربية في القرآن الكريم. يحتوي الفصل الخامس على المنهجية والأساليب والتجارب. وقد أجريت ست مجموعات من التجارب. أول واحد يستخدم أساسيات معالجة اللغات الطبيعية في نظام الرد السؤال بإزالة كلمات التوقف، والشكل، والرموز الخاصة. يستخدم الثاني الفهرسة في حزمة لوسين Lucene ونمط تفاعيل و تفاعيل، وفعاعيل، وفعاعيل. والرابع يستخدم في بناء ذخيرة لغوية ديناميكية لجمع الأسئلة من المستخدم الحقيقي. يستخدم الخامس نمط الصيغ المبالغة. يستخدم السادس نمط المفرد، والمثنى والجمع. وأخيرا هناك نتائج وتحليلها. وأظهرت التجارب أن إزالة كلمات التوقف و الشكل عززت

نتائج البحث. كما أضافت الأنماط الجديدة المزيد من القيمة لنظم الرد الآلية وتعزيز الاستدعاء والدقة.

Table of Contents

Introductory	2
Dedication	3
Acknowledgement	4
Abstract	5
المستخلص	6
Table of Contents	8
Table of tables	12
Table of figures	13
List of Abbreviations	15
CHAPTER I	16
1 Chapter one Introduction	17
1.1 Background	17
1.2 Problem statement	17
1.3 Objectives	18
1.4 Scope	18
1.5 Arguments and Motivations	19
1.6 Research Methodology	19
1.6.1 Tools: software and hardware	19
1.7 Thesis layout	19
1.8 Contributions	20
CHAPTER II	21
2 Chapter two Background and Related Work	22
2.1 Natural language processing	22
2.1.1 Terms	22

2.1.2	Natural language processing and techniques	24
2.2	An introduction to question answering system	25
2.2.1	History	25
2.2.2	Question answering (QA) system Definitions	27
2.2.3	Question answering system Components	28
2.2.4	Question classifier (Analysis)	29
2.3	<i>Question answering system domains</i>	33
2.3.1	Open domains	33
2.3.2	Closed domains	34
2.4	<i>Question answering system frameworks and architectures</i>	34
2.5	Corpus:	36
2.5.1	History and features	36
2.6	Chapter summarization	37
	CHAPTER III	38
3	Chapter three: related work	39
3.1	General question answering systems	39
3.2	Arabic question answering system	41
3.3	Question answering system and holy Quran computation studies	43
3.4	Chapter summarization	45
	CHAPTER IV	46
4	Chapter four Arabic language and holy Quran	47
4.1	Introduction to Arabic language	47
4.2	Arabic language characteristics	47
4.2.1	Arabic language general characteristics	48
4.2.2	Arabic language grammarian characteristics	49
4.3	Arabic natural language processing (ANLP)	50

4.4	Arabic language challenges	50
4.5	Arabic language ambiguity	51
4.5.1	Arabic language morphological ambiguity	52
4.5.2	Arabic language syntactical ambiguity	52
4.5.3	Arabic language semantic ambiguity	53
4.5.4	Some of the proposed solutions to resolve the problem of linguistic ambiguity in the Arabic language	53
4.5.5	Arabic language tools	54
4.6	Holy Quran	55
4.6.1	Introduction	55
4.6.2	The role played by the Holy Quran to save Arabic language	55
4.6.3	Holy Quran sciences	56
4.6.4	Holy Quran applications	57
4.7	Chapter summarization	57
	CHAPTER V	58
5	Chapter five: The experimental design and results	59
5.1	Tools and methodology	59
5.1.1	Java programming language	59
5.1.2	Microsoft Excel	60
5.1.3	Microsoft Access	60
5.1.4	Lucene search engine	60
5.1.5	Diagram designer	60
5.2	Building data set for holy Quran questions and answers	60
5.2.1	Procedures of collecting data:	61
5.2.2	Types of data:	61
5.2.3	Clean the data:	61

5.2.4	Details of building corpus:	62
5.3	Proposed Arabic question answering system conceptual framework model	65
5.3.1	The natural language processes in the model	65
5.3.2	The process steps in the model	66
5.3.3	Framework characteristics	68
5.3.4	Drawbacks of framework	68
5.4	Building prototypes for question answering system	68
5.4.1	The First prototype	68
5.4.2	The second prototype	74
5.4.3	The third prototype	77
5.4.4	Some important notices and observations for the three prototypes	80
5.4.5	Experiments and results	81
5.4.6	The fourth prototype (Enhanced prototype1)	84
5.4.7	The fifth prototype (Enhanced prototype1 sing exaggeration formula)	91
5.4.8	Prototype 6 singular, dual, and plural	95
5.4.9	Experiments Summary	106
5.5	Chapter summarization	108
	CHAPTER VI	109
6	Chapter six: Conclusion and recommendation	110
6.1	Conclusions	110
6.2	Recommendations for further research	111
7	References	112
8	Appendices	120
8.1	Some examples of the result of applying corpus in question answering systems	120

Table of tables

Table 5—1: Quran QA spreadsheet header	62
Table 5—2: Quran QA sample after expert validation	62
Table 5—3: Quran QA sample from WWW	63
Table 5—4: Differentiate between sources	63
Table 5—5: repeated ansers and questions from many resources	63
Table 5—6: double repeated question from different sources	64
Table 57—: Three repeated question from different resources	64
Table 5—8: the last table for corpus.	65
Table 5—9: Table of first comparisons	73
Table 5—10: Example of indexing some words by indexing sytem	74
Table 5—11: General table results for the three prototypes	81
Table 5—12:Percentages of right and wrong answers	83
Table 5—13: Displays percentages of right and wrong answer in matching	84
Table 5—14: comparison between prototype one and five	95
Table 5—15: the result word (patterns) of the query and some remarks	100
Table 5—16: General table results	103
Table 5—19: Prototypes summarizations	106
Table 5—20: Prototypes summarizations	107

Table of figures

Figure 2—1: Ggeneral architerture for question answering system.....	28
Figure 5—1: The proposed framework for Arabic question answering system.	67
Figure 52—: main processes of prototype 1	69
Figure 5—3: Entering the question form	70
Figure 5—4:results without removing diacritics and stop words.....	70
Figure 5—5: The question input.....	71
Figure 5—6: The relevant answer to the above question	71
Figure 5—7: Another input question	71
Figure 5—8: The answer after removing stop words and diacritics	72
Figure 5—9: the last question.....	72
Figure 5—10: the result after removing diacritics and stop words and punctuation.....	72
Figure 5—11: The indexing of holy Quran verses (Al-Baqarah and Alfatihah)	75
Figure 5—12: second and third prototypes processes	75
Figure 5—13:Inserting query in second prototype.....	76
Figure 5—14: The main words for the query after stemming some letters.....	76
Figure 5—15:The result of the query in figure 11.....	77
Figure 5—16: entering the query "هل للرياح تصريف"	77
Figure 5—17: The first prototype result for the query "هل للرياح تصريف"	78
Figure 5—18: The second prototype result for the query "هل للرياح تصريف"	78
Figure 5—19:The result of question for the word تصاريف.....	78
Figure 5—20:The result of question for the word تصاريف.....	79
Figure 5—21: Displaying main query words	79
Figure 5—22: The derivatives morphological of the word تصاريف.....	80
Figure 5—23: The result of question for the word تصاريف and their derivatives morphological	80

Figure 5—24:Displays percentages of right and wrong answer for prototype1	83
Figure 5—25: Displays percentages of right and wrong answers for prototype2 and 3	84
Figure 5—26: depicts prototypes 4 processes.....	85
Figure 5—27: The system asked the user to enter its question.	87
Figure 5—28:Accepting user query from the system.	87
Figure 5—29:The result of the question in figure two.....	87
Figure 5—30:The word الرحمن as its appeared in the file.....	88
Figure 5—31:The answer of the above questions	89
Figure 5—32:The changing of the word الرحمن in the text file.....	89
Figure 5—33: The result of the question after changing the word	89
Figure 5—34: The different answers displayed by another system.	90
Figure 5—35: An example of automatic corpus for Quran questions	90
Figure 5—36: Answers displayed by prototype 5	94
Figure 5—37: Noun category in Arabic language	96
Figure 5—38: user query من هو المؤمن.....	99
Figure 5—39:answer to the user query من هو المؤمن	99
Figure 5—40: the answer of the question من هو المؤمن	102
Figure 5—41: the answer of the question من هم المؤمنین.....	102
Figure 5—42:Comparative table for results.....	105
Figure 5—43: comparison between prototype 1 and 6.....	106

List of Abbreviations

IDE	Integrated Development Environment
ANSI	American National Standards Institute
EBCDIC	Extended Binary Coded Decimal Interchange Code
NLP	Natural Language Processing
POS	Part Of Speech Tagging
QA	Question Answering
IR	Information Retrieval
IE	Information Extraction
AI	Artificial Intelligence
NER	Named Entity Recognition
CR	Concept Relational
ABI	Attribute-Based Inference Abi
QIRS	Quick Information Retrieval System
CBR	Case-Based Reasoning
MSA	Modern Standard Arabic
MADA	Morphological Analysis And Disambiguation For Arabic
SVM	Support Vector Machines
VSO	Verb, Subject, Object
WN	Wordnet

CHAPTER I

1 Chapter one Introduction

1.1 Background

Internet data, increasing rapidly and growth in the last decade. Search engine and informational retrieval systems retrieve all documents that relevant to the query, this process depending on matching of keywords in documents and then retrieve the corresponding documents. These retrieved documents may not contain the required information, although if it found it may be in an unknown position in the document. One of the techniques to solve this problem is question answering systems. A Question answering system is a process of entering a question and processes it to find an exact answer for the question. Many processes were done to reach the exact answer explained through this research.

Arabic language which is the main language of the Holy Quran, also it took attention in the recent decades, so there would be a breakdown for the computerization of the Arabic language and the Holy Quran will be among the folds of this research.

1.2 Problem statement

This PhD project investigates the applicability of Question-Answering methods from Natural Language Processing research to answering Islamic users' questions, supplying answers from the Quran. We will review Question-Answering systems developed by NLP researchers for a range of other domains, including education[1] , biology and medicine [2], linguistics [3], and general answers from the Web [4] . In general, Question-Answering systems share three components: (1) Question Classification, as different types of question (definition, factoid, list, yes/no etc) require different processing [5]; (2) Information Retrieval, to fetch and rank the documents that are relevant to the question; and (3) Answer Extraction, to extract the required definition, fact, or other information from the retrieved documents. To apply this generic architecture to Question-Answering from the Quran, we treat each verse as a separate document or record; hence, the Information Retrieval stage aims to find and rank Quran verses which have the best match with the users' question. Given that each "document" or target Quran verse in this case is a short text, the third

stage of Answer Extraction may not be applicable or appropriate: for many users, the “answer” they want is just the relevant verse (or verses), for the reader to interpret themselves. For some users it may be helpful to extract and highlight the part(s) of the verse most pertinent to the question; but some scholars hold that it is not appropriate to “extract” and reformulate the words of Allah into another sentence.

To evaluate the applicability of different Question-Answering methods for answering Islamic users’ questions, the research will include collating a “gold standard” corpus of Islamic questions asked by real users, and answered by Islamic scholars. We compare Q-A systems based on stemming and keyword-match (e.g, [6], [7]) against Q-A systems using Key-word matching and removing stopping words . Our evaluation will show us whether, for the task of finding Quran verses relevant to the question without “answer extraction”, stemming and keyword-matching methods suffice; and whether there is significant added value from more sophisticated key-words matching and removing stopping words architectures.

1.3 Objectives

Our overall aim is: to investigate the applicability of Question-Answering methods from Natural Language Processing research to answering Islamic users’ questions, supplying answers from the holy Quran.

This breaks down into the following objectives:

- To adapt a question answering system based on stemming and keyword-match to holy Quran-Q-A-1.
- To adapt an alternative question answering system using key-words matching and removing stopping words to holy Quran-Q-A-2.
- To collate a “gold standard” corpus of Islamic questions asked by real users, with answers -relevant Quran verse or verses - as judged by Islamic scholars.
- To comparatively evaluate the two different approaches to Quran-Q-A.
- To build a question answering system framework for Arabic and Islamic application.

1.4 Scope

The research investigates Question-Answering research and systems, a broad research field; but will focus on how the general methods adapt to the specialized

task of Question-Answering from the Quran. As shown below, the overall research project breaks down into more specific objectives and work-packages, each of which is challenging but not insurmountable.

1.5 Arguments and Motivations

A clear motivation is to apply an academic research theory on Question-Answering to a useful domain with many potential users. Quran websites attract millions of users, and a Quran website augmented with an automated Question-Answering system could become very popular and useful. The research is also academically valuable, publishable in journals such as Applied Ontology Journal or Applied Artificial Intelligence Journal or International Journal of Computational Linguistics and Applications.

1.6 Research Methodology

1.6.1 Tools: software and hardware

1.6.1.1 Software: The software which is used in this project

- Java .
- Lucene search engine package.
- NetBeans IDE 8.5.2.
- Microsoft word.
- Microsoft soft Excel for some statistical calculations.
- Diagram designer.

1.6.1.2 Hardware requirement

- PC computers
- Network and internet connection.

1.7 Thesis layout

The thesis consists of six chapters, structured as follows: The first chapter explains general background, Chapter two general introduction to terms and concepts, chapter three related work, chapter four Arabic language and Holy Quran, chapter five is for experiments and results, chapter six is for conclusion and recommendations.

1.8 Contributions

1. Building Corpus from question and their answers from two Holy Quran chapters (Albagarah and Elfatihah).
2. Designing a general algorithm for Holy Quran question answering system .
3. Developing Six prototypes for Holy Quran question answering system .
4. Designing a general conceptual framework for the Arabic Question answering system.

CHAPTER II

2 Chapter two Background and Related Work

Natural language processing is one of the important computer science branches, which deals with languages understanding and generation; one of the important applications is question answering systems which apply many NLP techniques. For this this chapter contains more details and discussions about NLP and question answering system. This part explains general concepts , some computational linguistical terms, and their classifications. Also in question answering system history take place, its components, techniques used and domains.

2.1 Natural language processing

It is a branch of computer science that concerns with generating and understanding natural language processing, this includes the three main skills for language which are: reading, writing and speaking. In general, computer see the text but it doesnot understand it and need some knowledge to reason, for this many processes were done depending on many linguistics theorems and many tools. In this part the researchers review this area in more details.

2.1.1 Terms

In this part the researchers will discuss some important terms in language and then discusses how can computer deal with these terms and their usefulls in computation.

- **Synonymy:** different words that have the same meanings.

Different: diverse, unlike.

Refer: mention, state.

- **Antonyms:** words have opposite meaning.

High: low.

Big: small.

- **Hypernyms** : words that refer to a general category and a specific instance of that category.

Color is a hypernyms of red.

- **hyponyms:** is the inverse of hypernym.

Red is a hyponyms of color.

- **Polysemy:** poly means many and semy means sense in general a word that has two meaning or more such as the word get and set.

Accept: agree, admit.

- **Homophones:** are words that have different meanings, but the same pronunciation.

Fisher: fissure.

- **Homographs:** words that have different pronunciations, but have the same spelling.

The plural of axis, axes - the plural of ax or axe

Frequent: visit a place with regularity, occurring regularly to

- **Homonyms:** are words that have the same pronunciation or spelling, but have different meanings.

Meat: Meet.

Bee: Be.

In computation linguistics these terms are very important, because we can build many relationships between words to use it in the semantic process after building many rules and laws. Many tools can be built depending on these relations in computational languages, one of these tools is Wordnet. These relations can be used with words only, also an application can apply this is: information retrieval and question answering system query expansion. Another term used in a word sense is:

- **Synset:** which is a semantic relation between word senses.
- **Patterns:** defined by Cambridge English dictionary² as: " a particular way in which something is done, is organized, or happens" and some times called template.
- **Corpus:** defined by Cambridge English dictionary³ as : "a collection of written or spoken material stored on a computer and used to find out how language is used ".

² <http://dictionary.cambridge.org/dictionary/english/pattern>

³ <http://dictionary.cambridge.org/dictionary/english/pattern>

2.1.2 Natural language processing and techniques

Also, many techniques are common in language computation such as: stemmer and tokenizer. In the following part the researchers discuss them in more details.

- **Removing:** is the process of removing some texts that not belong to the main text such as code, some symbols such as ?,. and ", removing diacritics.
- **Encoding:** is the process of changing the type of encoding character from system to another, such as: ANSI to EBCDIC and EBCDIC to Unicode. It used to display characters in the right way.
- **Tokenizing.** : is the process of dividing the sentences to words depending on the sentences separator which is vary from sentence to another and the boundary of the words of the sentence. It is defined by [8] as " Tokenization is the task of cutting a string into identifiable linguistic units that constitute a piece of language data" . Many separators can be used such as: space, , , -,etc.
- **Stemming:** is the process of removing some additional characters from the word to categorize a group of words into one main category to help in semantic analysis. It is defined by [9] as "is the task of correlating several words onto one base form"
- **Morphological analysis:** is the process of dealing with relationship of the words with the same inflection (structure of the word depending on its position and its classification in the sentence).
- **Part of speech tagging, POS tagging:** is the process of "Automatically assigning parts-of-speech to words in text is" [8], which hepls in finding accuarte answers in question answering system[10].
- **Segmentation:** is the process of splitting the text in to meaningful parts such as: words, sentences.
- **Capitalization:** is the process of writing the first character in the word with capaital letter. This characteristic is very important in named entity recognition NER to identify the right answers.

2.2 An introduction to question answering system

Question answering (QA) system is a big research area in many fields and domains, which can help users in specific or general domains to find exact and precise short answers if found, instead of returning a ranked documents that contains phrase or word(s) in phrase matches any words of documents in information retrieval. Many fields of science intersect with QA such as information retrieval (IR), information extraction (IE), Natural language processing (NLP), artificial intelligence (AI), and computational linguistics.

[2] Explain that the main aim of QA is to provide users with exact and precise answers by using Information extraction (IE) and NLP techniques.

[11] Explains that in question answering systems we must consider the following factors: applications, users, question types, answer type, evaluation, and presentation.

2.2.1 History

In this part the researchers need to revise the history of question answering system from the earlier time till now. What is the technologies was used, methods and applications and other things.

[12] proposed Chatter bot which is a program that simulate conversation between people and machine (robot answer the user question) in Tamil language. The system initiate the conversation user can write its query (from simple to definition questions to complex) in any way then the system which is called POON GKUZHALI reformulates the question and gives the answers. Each question is processed separately, also can be references for the other question in knowledge base.

The first proposal in this area written by Simmons, R. F., "Answering English Questions by Computer: A Survey," [13].

One of the first question answering system is the BASEBALL: AN AUTOMATIC QUESTION-ANSWERER which is used to answer the question about base Ball in America for one season. The system accepts the question in English phrase and then compares it with data stored in punched cards if the answer is extracted then the system answer the question. The system has a four NLP

routines which are: question read-in, dictionary look-up, syntactic analysis, and content analysis [14].

[15] explained that many question answering system described and revised in translation and coversation. These systems were classified as inferential, text based, and data basd. Also, he explained the difficulties facing peoples at that time such as: measuring, ambiguities, translating , and searching .

[16] represented that most of the question answering system in that time developed for syntactic, semantic, and logical analysis for texts. They used small subset of data, but the main problems was the generality of questions, grammar, and semantics.

[17] explained the semantic models in that time were based models of human cognitive processes and were applied to a small dataset of English. Most of these models are tentavives because semantic processes are not drafted rigorously.

[13] Nasa Manned Spacecraft Center supported Lunar as research prototype for geologist to easy access, analysis, and evaluate chemical data on Lunar rock and soil that sent by Appolo moon mission. The query put in formal query language and then use some predicate calculus, then manipulated to drive inferences or access database directly to drive other inferences.

[18] proposed an English language question answering system for a large relational database of aircraft flights (PLANES system). PLANES uses augmented transition network for matching phrases with a specific meaning compared with history keeper and concepts case frames; These used for judging the significance of the questions. Also PLANES used dialogue to deal the question partially to clarify the question and the system solve the problem of ellipsis and pronoun reference. The system accepts the natural language question if it is wrong in grammar,then the system returns explicit answers in form of text, number, and graph if possible. Also the system is tolerant with grammar and spelling errors,online, interactive, easy, and gives about 75% and 90% and more right percentages.

[19] proposed QUEST which is a question answering system to simulate answers that produced by human in open class question(when,who,etc.) and closed class question(is y true or false). The system identifies the relevant information sources, these sources organized in graph structure. It used artificial intelligence and

computational linguistics concepts, also the system intended to be psychological of question answering.

[20] proposed MURAX which is used robust linguistic methods are used in answering closed class questions using a corpus to answer general question using an online encyclopedia. The methods inspect the answer is a noun phrase and return the relevant texts to the users. These matching text used with question to assure phrase relation implied by the question. The query build automatically and then passed to the IR to find relevant result which is analyzed. Noun phrase extracted as result and new relation for various hypotheses.

[21] proposed QACID which is an ontology-based question answering system. It is accepted the user queries in natural language then it retrieves the information from formal ontologies. The system applied in cinema domain. The approach it depends on collections of user queries, also the system present simple adaptation to deal with multilingual capabilities, inter-domain portability, and changes in user information requirements.

[22] proposed VNEWSQA/ICT which is question answering system deal with questions that accepted by the system then use two models one to deal with Vietnamese language as natural language processing and the other deal with Vietnamese semantics. The system extract the answer extracted from titles of Vietnamese news in the domain of information and communication technology.

We notice the question answering systems ranging from simple text question answering systems with the earliest problems of measuring, ambiguities, translating, and searching to advanced question answering system that used semantics and syntactic models specially, and from specific domains to general domains specially that systems which used English language.

2.2.2 Question answering (QA) system Definitions

Question answering (QA) system is a specialized type of information retrieval that returns precise short answers to queries posed as natural language questions [5]. Also [23] defined it as Question Answering (QA) Systems allow the user to ask questions in a natural language and obtain an exact answer. [24] defined it as: Question answering (QA) system is a specialized area in the field (IR) of

information retrieval which is concerned with providing relevant answers in response to question proposed in natural language processing. I can define it as a system that accept a question from users in natural language then processes it by applying natural language processing techniques to obtain an exact and short answers instead of retrieving full documents as result of the users questions if found by applying many AI and linguistics methods.

2.2.3 Question answering system Components

In general, Question answering system has three main components: Question classifier, Information retrieval (IR) and answer extraction. Each one of these three components can be applied as a module with its characteristics that it can vary from system to another, domain to another, language to another. . Figure[1] below explains the general main components of question answering system.

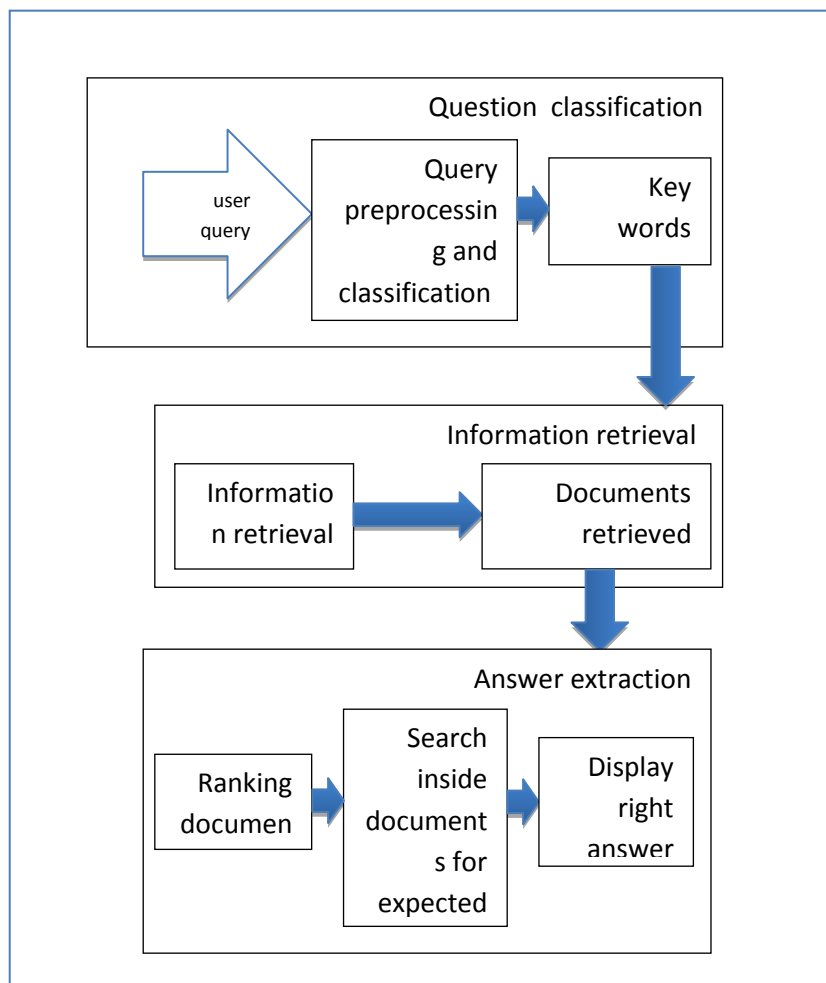


Figure 2—1: Ggeneral architerture for question answering system

2.2.4 Question classifier (Analysis)

[25] determine that: there are three types of methods that used in classifying texts those are: template based approach, machine learning approaches and knowledge based approach.

There are many different types of questions, used to identify the answers (two categories of questions in Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A) which is guide to build QA system).

The first taxonomy is: Arthur Graesser's Taxonomy of Inquiries which contains eighteen questions: Verification, Comparison, Disjunctive, Concept completion, Definition, Example, Interpretation, Feature specification, Quantification, Causal antecedent, Causal, Consequence, Goal orientation, Enablement, Instrumental /Procedural, Expectational, Judgmental, Assertion, Request / Directive. The second category is : The 13 conceptual question categories used in Wendy Lehnert's QUALM : Causal Antecedent, Goal Orientation, Enablement, Causal Consequent, Verification, Disjunctive, Instrumental /Procedural, Concept Completion, Expectational, Judgmental, Quantification, Feature Specification, Request.

[26] explains that Arabic is one of the most rich and more complex , morphologically, also Qur'anic Arabic is more complex, for this we need to applying morphological analysis to identify structure such as: morphemes, affixes, roots, part of speech, stem, etc.

There are many preprocessing methods and techniques used in analyzing text such as:

2.2.4.1.1 Tokenizing the question

Lexical analysis is the process of converting a sequence of characters into a sequence of tokens. Tokenizing the question means divide the question phrase entered by users to words to use each one in matching with other words in the database or knowledge base or information resources for the system. We can use the

words as it is instead of using any other linguistics techniques in search but this decrease the answers e.g. WordNet.

2.2.4.1.2 Stemming words after tokenizing

[27] defined it as method of word standardization for matching terms morphological related. In Arabic language, we add affixes: (prefix, infix, and suffix) to modify the meaning of the word. By stemming word the matching words in the result may be increased by finding many words have the same meaning.

2.2.4.1.3 Parser

The parser is used to find out the grammatical structure of sentences and sometime tagging words as in Stanford parser. Also, there is a chunking which is a partial parsing which is dealt with partial syntactic structure.

2.2.4.1.4 Part of speech (POS) tagging

Part of speech is to put words into classes or lexical categories. Eight types of POS which are noun, verb, adjective, adverb, pronoun, preposition, conjunction, interjection, and sometimes numeral, article or determiner.

The tag is grammarian agree, where part of speech tagger is to processes a sequences of words to find the corresponding POS to each word-tagging can be done manually or by machine or both to disambiguate word structure or word category. We use tagging with systems that check the syntax.

2.2.4.1.5 Ontology

[25] defined ontology as a technology used to enable the domain knowledge at the high level and improve the query time used in question answering system. We use an ontology with systems that check the semantics of the sentences and words.

Many other features used in question analysis, such as: Sentence boundary detection which it concerns with where the sentence begins or ends, Sentence pattern detection many structure used to define these sentences such as 9/1/2015 6:14:10 PM for date and time, Named entity detection. Also in this part, we need to remove: stop word (common words in a language that has less value in research), diacritics to enhance the word matching in this stage.

All the analysis and processes to formulate and rephrased the query entered by users, to find accurate and suitable answers.

In this part some important techniques to expand the query by adding new key words to the main key words of the query after making the important preprocessing. These techniques are: language resources (lexicon[28],dictionary[29] and thesaurus[9]), ontology[25], stemmer (Khoja[30] stemmer which abstract the word to its root and Arabic light stemmer [31] which remove some prefixes and affixes) and patterns[32].

At last we can summarize this stage in three main processing: Analyzing the query to find the main key words to answer the question, classification the question to identify the type of question, reformulate the query to enhance it and to pass it to the IR system.

2.2.4.2 Information retrieval (IR)

From this[11] said that :this part has been active research area since mid 1950s. [33] explained that standard IR is: " Information retrieval (IR) has most usually been constructed as the problem of selecting texts from a database in response to some more-or-less well-specified query". Also [34] define it as : " Information Retrieval (IR) is the discipline that deals with retrieval of unstructured data, especially textual documents, in response to a query, which may itself be unstructured like sentence or structured like Boolean expression". Search depends on query users and full text search, meta data,indexing, etc.

The goal of IR is to retrieve accurate documents contain results and ranked[35] depending on users queries. In IR if the key words matching any words, lexical or semantically in the document resources (structured, semi structured, unstructured) depending on many methods such as: Boolean, probabilistic, vector space and semantic this leads to a good performance. Also IR has the following features: matching partially or full, induction inference, probabilistic model, the source of data is documents, the query may be in natural language or not, matching may be relevant,etc.

When the query passed to the IR system, the system search in one corpus or many corpora to find the relevant document(s), that matches the phrase or key words

in the query then the IR system must return Ranked the documents depending on many criteria's such as : matching full phrase, matching key words, number of key words matches,etc and filtering the documents to find the repeated documents to decrease the number of candidate documents.

The main problems of dealing with IR systems in Arabic stated by [26] in three main points as follows:

1. The search engine retrieves the relevant documents but not the exact answer, this means you need to search inside the documents for answers.
2. The quality of data because users of social media explains their opinions.
3. Challenges of Arabic language.

2.2.4.3 Answer Extraction

In information retrieval the search engine returns the documents that contains the keys or terms in query after preprocessing the query, then you need to search inside the documents because their main aim isto retrieve only documents [36]. But in information extraction there is a deep processes inside the documents to extract structured texts from un structured texts[37].

The answer depends on the application answer source which were divided into three categories : structured data (databases), semi-structured data (for example, comment fields in databases) or free text [11].

In this part there are three main processes which must be done: first when the IR returned the relevant documents after identifying the type of answers by parser such as Named Entity Recognition (NER) and taggers(here some answers are identified directly in this stage). Second part is used ranked candidate answers by the previous steps then the system uses a set of heuristics to extract the relevant word or phrase also depending on many criteria such as full phrase matching, number of key words in the result, distance between number of key words etc. Then the candidate answers returned depends on many measurements such as probabilities of number matched words between query and answers, the type of question and its expected answers, the relationship between Is-A and its answers in a definition questions, the maximum count of NE(Named Entity) occurred in the question and

candidate answers, the distance between matched terms in passage[38], these measurements are for Name Entity Recognition.

[39] explains that many criteria's can be decided about the right answers based on a set of characteristics that describes:

- The interoperability between question and answers types.
- The redundancy of answers across the stream.
- The overlap and non-overlap information between question-answers pair and the supporting text.

Also [40] explained many features to evaluate answers, these features are:

- Relevance: we need to find the relevance between documents in the system.
- Correctness: the answer must be right.
- Conciseness: we need to count the number of byte answers and remove unimportant part of the answers.
- Completeness: the answer must be complete, if we had to use additional sources.
- Coherence: an answer must be consistent.
- Justification: the answer must be supplied with sufficient inference tools to determine why the answer had been chosen.

2.3 *Question answering system domains*

In general question answering system classified into two main domains: open domain and closed domain [41].

2.3.1 Open domains

Question answering system open domain deals with general questions about anything rely on universal ontology and information such as web. Many question answering system proposed in the domain such as : [42] proposed Processing Definition Questions in an Open-Domain Question Answering System, [43] proposed Designing an interactive open-domain question answering system, and [44]proposed Domain Ontology Construction using Automatic Concept Relation Learning for Question Answering Systems.

2.3.2 Closed domains

Closed domains answering systems deals restricted and specific domain (cinema, health, sport) Many researches in many areas were done such as:

- Health :[45] proposed Creation and use of language resources in a question-answering e-health system.
- Education : [46] proposed A Hindi Question Answering system for E-learning documents, and [47] proposed BionQA Multi document Question Answering System: Providing Access to E-learning for masses.
- Biomedicine : [48] presented Multilingual Question/Answering System applied to trusted health information, and [5] proposed Usability survey of biomedical question answering systems.
- Linguistic: [3] proposed: A Proposed Online Approach of English and Punjabi Question Answering, and [23] A prototype framework for a Bangla question answering system using translation based on transliteration and table look-up as an interface for the medical domain.

2.4 *Question answering system frameworks and architectures*

Many architectures proposed in question answering system, which is vary in its components and shapes, but in general, most of these architectures Agree on its basic components.

Core question answering architectures proposed by [43], this architecture consist of the many parts question, question classification, web retrieval, web documents, documents processing, factoid question to determine the type of questions if factoid or no factoid to use suitable answer extraction for the factoid or nonfactoid answers.

An architecture of base line serial system with no feed back proposed by [49], this system has the following processing: keyword preprocessing (split, bind and spell), consruction of question representation, derivation of expected answer type, keyword selection, keyword expansion, retrieval of documents and passages, identification of candidate answers, answer ranking and answer formulation.

A generic architecture for question answering system proposed by [11] which is depending on requests from users and responses from the system. The users enter question, this question analyzed by user model, dialogue context and response generation, then send two messages to candidate document selection to select candidate documents and answer extraction. Then the candidate documents selections send these documents to answer extraction to select right answers and ranked them.

An other intelligent system architecture proposed by [50] which is PIQUANT's system architecture. The system consist of four part question analysis, answering agents, knowledge sources, and answering resolution. The search use two methods: semantics and keyword search. Answering agents consist of four types: knowledge based, statistical, definition and CYC based(CYC: a large-scale investment in knowledge infrastructure [51]). The system also uses many resources such a corpus, database and ontology.

Also PiQASso designed by [52], the system accepts the question and parses it through MiniPar parser, then the system classifies question and expands it by using WordNet. The system search paragraph in documents to search for the relevant paragraph with matching relation and type, also using WordNet. The last steps to score and rank answering to display the answer.

A question answering architecture for Hindi proposed by [46], the system classifies user question depending on some rules and parsing it. After this the system formulates the query to use their words for matching answers from corpus, then the system retrieves passages after scoring and ranking.

A semantic processing model for Vietnamese architecture designed by [22], the system use a Vietnamese dictionary to determine question words and their syntactic categories, the grammar rules used to determine phrase, analyze and syntactic structure, by using semantic representation rules to determine the semantic representation of the question, lastly, there is a matching semantic representation of the question with semantic representation of the database.

A Quranic question answering system architecture proposed by [26]. The system has two main parts: Online and Offline. The first part for question analysis (classification and processing), information retrieval to retrieve verses that containing answers and answer extractin which containg (Arabic NER and extract answer). The second parts containing Quranic ontology, Off line processing for Quran verses to weight vector of concepts of each verse. The system merge the two parts to extract the right answer.

2.5 Corpus:

This term until now there is no agreement to it[53], but we can use this definition which is: A corpus is “A collection of naturally occurring language text, chosen to characterize a state or variety of a language ” [54]. The singular is a corpora. Many and large texts written and spoken, these texts can be used for more statistical and researches to reach to good informations and knowledges. These texts can be collected from many resources such as: conversations, perliamints documents, medicine records, news papers etc. It can be use one language or two or more which is called comparable corpora[53].

2.5.1 History and features

The first corpus created in early 1960s designed by Nelson Francis and Henry Kucěra, which is Brown Corpus. It is a standard corpus of American English which consisting of 1,014,312 words.

It comes from the Latin word, which means the body. A corpus (singular is corpora) is a linguistics approaches the study of language in use as expressed in examples through corpora. It is a collection of examples in real world texts that used in analysis of bodies corpora (large bodies of data) to produce a large set of data suitable for computer processing such as: which pattern is associated with grammatical and lexical features and how these patterns are varieties within sentences and registers

A corpus gives users good features for languages such as: frequent words and phrases in languages, differences between spoken and written languages, use of

tenses and prepositions, uses of phrasal verbs, use of words in formal and informal situation, uses of words in conversations, etc.

2.6 Chapter summarization

This chapter investigates natural language processing in general, some term and its techniques. Also question answering system explained and their components, domains and frameworks. Finally a corpus introduced with its features and history.

CHAPTER III

3 Chapter three: related work

There are many researches were done in this area. We will divide this part to three parts, which are: general question answering systems, Arabic computation and corpus, Arabic question answering systems and Islamic computation and question answering systems for holy Quran.

3.1 General question answering systems

Many question answering systems were proposed from early sixties, but we will concentrate our studies in some important system and see some of its characteristics.

[55] proposed REQUEST which is question answering system implemented in lisp, it had the ability of analyzing and answering many types of questions and deal with 500 of databases. The goal of the request is to build an interactive machine that communicates between people and computer. He thought about the conflicting requirements between human and machine, and taking into consideration three important characteristics: the use of restricted English; Linguistic analysis based on a transformational grammar; and a two phases, compiler like organization. Its goal is to cover limited scope of semantics. The system has more than one hundred transformational rules for grammar. One of the draw backs of Request: can't be applied in general domains, limited coverage of questions, it covers small database from 1967 to 1973.

[56] proposed a system uses a clue words that used to defined the suitable topic from all topics that used to retrieve the documents and interacts with user to explain are these clue words are suitable for the reformulated query. The system modifies the clue words in each query as a constraint then these words classify to groups for Named Entity Recognition. The system uses these clue words to expand the search space and retrieve documents to increase possibility of getting answers candidate, and then displays the expected documents that contains answers, then interact with users to determine the appropriate answer. This stage repeated until the user reach to the right answers. Although the system use a small topics of questions, sentence generation method for interaction is need to improve.

[57] proposed approach to identify the word sense in the sentences. They explain that word sense disambiguation can be seen as a classifier for context and classes to provide evidence. Each occurrence of the word can be evidence, and these occurrences can give finite sets of senses from resources. Word sense disambiguation algorithms can be classified as corpus base approach and knowledge base approach. The system uses a database and the relevant relationships created by WordNet. Then, designing an algorithm to serve many and wide groups of words in a specified domain. The evaluation of the algorithm compared with a heuristic approach shows that there is a higher precision-recall rate for the algorithm. However, the system has a high precision but it used only five words for its experiments which is not suitable for evaluating the system results.

[58] they used domain ontology as a source of knowledge for a question answering system. They state that the ontology can supply the term with many other meanings of the term if it is defined well. They explained that automated ontology solved many problems of building manual ontology such as: difficult and time-consuming task, the dynamicity, potential size, and complexity of a specific domain increase the difficulty of building manual ontology. The solution of these problems is to use automated ontology. They proposed a methodology to build an empty ontology to update one by extracting relations from unstructured data using Naïve Bayes classifier, which extended to learn concepts relations from extracted associations. Then the predicted relations were used to model the domain concepts for the resulting ontology. In addition, they proposed a framework for the concepts relational Ontology based for question answering system processes. The question answering system framework consists of a dynamic concept relational (CR) ontology construction module to extract new concepts from the web and then keep it in an ontology knowledge base, an answer extraction module.

Then they proposed a Concept Relational Ontology-based Question Answering system (CRO-QAs) that consists of a question analysis component, an answer extraction component, and an automatic ontology construction component. To extract the answer from the ontology they used Attribute-Based Inference (ABI), which was introduced by [21]. The system uses three types of questions: factoid, list, and definition. The relation extraction was evaluated using benchmark data sets. The experiment explained

that the system can be used to construct generic domain ontology with high accuracy. Then the system developed to question answering system framework. However, the system is more complex and use only three types of questions.

[59] explains answering system based on CBR (case-based reasoning), which is a problem-solving technology based on experience, this technology used historical and valid answers to answer new question. Also he illustrates the CBR advantages which are:

- Easy access to knowledge.
- Easy maintenance of the knowledge database.
- Does not need domain experts' interference.

The system applied in distance teaching platform. By applying CBR to the answering system, the system does the following:

- Classify and extract the questions.
- And classify and extract the answers which stored in the library.
- Build a case base that data complete, accurate, easy, organized retrieval and maintenance.

The main challenge of the system when a new question is entered, the system can improve itself by adding the question and its answers to the database.

The system used automatic segmentation, questions similarity calculation, improve search efficiency which they can improve search efficiency. The system uses the method of redundant storage the keyword to improve the speed. The results of experiments show that the system can improve the accuracy and intelligence of answering systems.

3.2 Arabic question answering system

[60] proposed an Arabic question answering system QARAB which is accepts the question in natural language (Arabic language text) and tries to find short answers to the questions under the following assumptions: the answer strings extracted from the Al-Raya newspaper published in Qatar and all the answers appear in one documents. The system has four main processes, which are: enter the question, matching the key words of the question with the database of answers, retrieve passages which contain candidate answers and displaying five ranked

answers with its document references. The system tagged the question words and add it to lexicon if it is not there. Then the system classified the question types to determine the type of answers. The system uses QIRS Quick Information Retrieval system, which is retrieved passages that matches the queries, then the system parsing these passages to recognize named entity to be possible first five answers.

Two experiments were done: question answering using the modified-word strategy (bag of words) and question answering using the root-based strategy (question expansion). These two experiments were judged by the users that testing the system by entering the question and then see the results are correct or not correct by using MRR (Mean Reciprocal Rank) give result about 86% of correctness in general. The system uses a simple type of question, but not used complex question such as what, why and how.

[61] proposed QASAL which is an Arabic question answering system used natural language techniques to handle Arabic factoid questions. QASAL consist of three components: Question Analysis, Passage Retrieval, and Answer Extraction. Then QASAL uses NooJ's platform, and developed to answer Definition questions (questions asked for person profile, full name of concepts, or an organization) by using three modifications: first, using of question patterns, second, using of the Google search engine and the Web as resources, and last using of a list of lexical patterns to extract answers. The system results obtained are recall equal to 100% and precision equal to 94%. However QASAL targeted only factoid and definition questions and uses lower of question numbers.

[28] introduced QAS which is an Arabic question answering system models that using Arabic grammar and morph-syntactic patterns. The NLP system has many modules which are Tokenizer, Tagging System (or type finder), Feature Finder, Proper Noun Phrase Parser. The IR models retrieves all relevant documents based on Salton's vector space model. The system uses dependency on data redundancy instead of complicated NLP analysis. The system is limited because the number of available tools developed for the Arabic language, also the system experimented by using twenty five documents and twelve queries.

[62] developes Yes/No Arabic question answering system that contains three question analysis module, text retrieval module and answer selection module.

The first module removes the following: the question mark, the interrogative particle, the stop words, the negation particles, and the other processes, which are tokenizing, tagging, and parsing. The system checks the question depending on nominal and verbal sentences, and then the system generates the required logical representation. After this check if the question negated or not, else reject it. In the verbal sentences, the system returns a list of synonyms and antonyms for the verb and in the nominal the system returned the root depending on the thesaurus provided by the Microsoft word Arabic supported version. The main idea of the system is logical representation, it used 12 Logical representations for the Nominal and verbal Sentences. These representations described based on the syntactic knowledge extracted from Arabic sentences that determine the structural role of certain words. For the semantic, the system uses certain word meanings and how these meanings joined in sentence representations forms. After these processing there are two types of information retrieval: document technique and paragraphs technique. Each document split into paragraphs and each one with its reference number and then tokenized, then it ranked by a mathematical formula. Also the documents use tokenizing and removing some addition like stopping word, then ranked documents depending on mathematical formulas. Last the yes or no answers generated after some processes. The system uses a corpus of 20 Arabic documents and a collection of 100 different yes/no question. The results comparing with answers generated manually they found that the paragraph techniques results better than documents. A percentage of error results are about 12% was due to the failure of the automatic tagging, the syntax of the user question, different words have the same root. However, the system uses a few numbers of documents and questions, and the system does not solve the problem of conditional responses.

3.3 Question answering system and holy Quran computation studies

There are many studies for holy Quran such as: [9] which provide a novel diacritic-less searching approach to retrieve relevant verses that match a user's query through automatic query expansion techniques. [63]proposed a system that retrieves any Qurnic verses and text and any other knowledge that related to holy Quran and its citation. Also [64]build QurAna which is a Corpus of the Quran annotated with

Pronominal Anaphora which characterized by large number of pronouns tagged with antecedent information, maintenance of an ontological concept list out of these antecedents and covering classical Arabic text. This corpus can be useful for applications in Modern Standard Arabic (MSA), will enable researchers to find patterns and rules to build new anaphora resolution approaches, and can be used to optimize, train, and evaluate existing approaches.

[65] proposed SPARQL an ontology semantic approach to extraction of knowledge from holy Quran. In this approach, the system reformulates the question to use it against the Quran ontology stored and annotated in the knowledge base. The ontology consists of noun concepts and its relationships. If the query matches the relevant answers, then return the answers which is verses.

However, there is a lack in question answering systems for holy Quran, but there is only one system, which is Albayan (according to my knowledge).

[26] proposed an Arabic question answering specially for holy Quran. The system has three main stages: input the question to make the preprocessing and analyzing the question to classify question. In the second stage the relevant verses retrieved semantically offline using Quranic data. Finally the answer extracted from the relevant verses and their Tafseer (the general meaning of this answer). Many preprocessing were done for the question and results such as tagging by POS, applying morphological analysis by MADA (Morphological Analysis and Disambiguation for Arabic), question classification done by rule-based approach; the system uses an SVM (support vector machines) classifier for training data from a corpus of questions and its relevant answers. Also, they proposed a new taxonomy for named entity recognition NER, which is a subtask of answer extraction to define concepts, names, and locations. Lastly, they constructed an Arabic Question Classifier. Evaluation results display that the overall system can reach until 85% accuracy, using the top three results. However, the classifier cannot determine the correct answer if the type of the question is of the second type (when we use question begins with the verb to be).

3.4 Chapter summarization

This chapter revised the related work for question answering system in three areas: general question answering system, Arabic question answering system and question answering system applied in the Holy Quran.

CHAPTER IV

4 Chapter four Arabic language and holy Quran

Now a days, natural language processing takes place in computation. Some of Latin languages take attention earlier and reach maturity, but Arabic language has taken attention in the last two decades for its importance. Also, Holy Quran which is written and cited in Arabic language also take place for its importance. This part of the thesis, explains Arabic language and its characteristics, challenges, ambiguity, Holy Quran, its role to save Arabic language, and its sciences.

4.1 Introduction to Arabic language

The Arabic language is a Semitic language[66], it is one of the official languages of the United Nations[67], depending on [68] it is ranked as fifth language, it is spoken and written by over 300 million person in the world[30]. It is used diacritics " the orthographic convention of leaving out short vowels and other distinguishing marks in written text"[69] which can change the meaning and pronunciation. Also it is the language of the pre-Islamic poet[70]. Also it (classic) doesn't change more than 50 century[71]. It can be classified into three categories[28]:

1. Classical Arabic: Classical Arabic or Quranic Arabic is the standard form of the language used in the holy Quran[72], the Scripture text for Muslims. In general it used diacritics and some symbols to make reading of holy Quran right and precise.
2. Modern Standard Arabic: it is more familiar and formal format of the language in all Arabic countries which is used in most conversations. It is used in education, media, newspapers etc.
3. Colloquial Arabic dialects: where each of the 27 Arab countries has their own dialect. Dialects are different from country to another and from region to another region in the same country[30].

4.2 Arabic language characteristics

The Arabic language has many characteristics, general and grammarian characteristics in this part we will discuss some characteristics depending on these two categories:

4.2.1 Arabic language general characteristics

Many studies were done, but [73] explains the following characteristics:

- (a) Arabic language written from right to left. .
- (b) It has three numbers: singular, plural and dual.
- (c) It has two genders: feminine and masculine.
- (d) Arabic words formed from roots which are triconsonantal and Quadriliteral roots.
- (e) Gender affects to nouns and adjectives.
- (f) The feminine nouns accept masculine numerals and vice versa.
- (g) Suffixes and prefixes refer to the complete actions in the verbal tense.
- (h) The suffix "ta" refers to grammatical gender "feminine".
- (i) Richness of vocabulary.

Other features explained by [74] which are:

- (j) There is no complete formal description of Arabic language.
- (k) Arabic language is inflectional language.
- (l) Arabic word can consist of particles and affixed pronouns.
- (m) Arabic high syntactical flexibility.

Also [75] states the following characteristics:

- (n) Arabic dialects vary depending on geography and social factors.
- (o) It is the language of Quran.

Other features are in Arabic script explained by [76] which are :

- (p) Cursivity: The Arabic word letters stuck with each other, this lead to increasing the number of forms of Arabic letter.
- (q) The shapes of Arabic letters changing from position to other(isolated, initial position, medial, and final.
- (r) Arabic character width is variable, in spite of they don't have capital or small letters.
- (s) Some Arabic letters that have diacritical points.
- (t) Hamza is used as vowels in Arabic language.
- (u) Arabic language uses diacritics to explain vowels.

4.2.2 Arabic language grammarian characteristics

In general, Arabic words classified into three main categories: Nouns, verbs and particles, each with its sub categories.

4.2.2.1 Nouns categories

Arabic nouns are divided into main categories such as:

- **Definite or indefinite** : The definite nouns categories are :
 1. The Proper Noun . العلم
 2. The Definite Noun with (AL ..) الاسم المعرف بال
 3. The Pronoun الضمير
 4. The Relative Nouns الأسماء الموصولة
 5. The Demonstrative Nouns أسماء الإشارة
 6. The Indefinite Noun added before a definite noun from the above. المضاف إلى معرفة
 7. The Noun Addressed by (Yaa) (يا) المنادى بـ
- **Gender** : nouns gender categories are:
 1. Masculine.
 2. Feminine.
- **Number**: nouns numbers categories are:
 1. Singular.
 2. Dual.
 3. Plural.
- **Derivation**: all nouns are either
 1. Not derived and nothing is derived from them,
 2. A source of derivation (also known as a gerund), or
 3. Derived from a gerund

4.2.2.2 Verbs categories

In general Arabic verbs based on a set of three, four and five characters as the roots of the words which is fixed and derived verbs. It has many categories, but the main category depends on time: past present and imperative. But when we use vowel we change the grammatical function: gender, number, person and voice. Also

can be classified as: genders (masculine, feminine), sound and weak, voice (active, passive), and numbers (singular, dual, plural).

4.2.2.3 Particles

We can divide particles into the following: preposition particles, the particles that resemble verbs, conjunctions, particles used for alerting, vocative particles, particles for affirmative answers, particles used for negative answers, extra, particles that introduce an explanatory sentence, gerundival particles, particles use for prodding, the particles used to indicate the nearness in time or certainty, conditional particles and interrogative particles.

4.3 Arabic natural language processing (ANLP)

Natural language processing for the English language started earlier and reach its maturity depending on the researches done in this area, but with the Arabic language it is different, in the last three decades Arabic natural language processing take place in Arab worlds and some universities and scientific institutions of Western. There are many motivations for Arabic natural language processing, [71] explains four motivations which are:

1. Transfer of knowledge and technology to the Arab World.
2. Modernize and fertilize the Arabic language.
3. Improve and modernize Arabic linguistics.
4. Make information retrieval, extraction, summarization, and translation available to the Arab user.

Also, there are other motivations such as:

5. Computation of Holy Quran and Sunnah.
6. Designing tools for Arabic natural language processing.
7. Designing a good Arabic search engine.

4.4 Arabic language challenges

Many challenges facing Arabic language. Many studies were done in this part, but one of the important studies that one done by [71] which states that: there are two types of problems: social and linguistics structure. These problems are:

1. The first problem was: the foreign names when it's written in Arabic there is no consistency such: "Washington could be spelled 'واشجنطن', 'واشنطن', 'واشجنطن', 'واشنطن', 'واشنطن'".
2. The lack of Arabic named entity recognitions.
3. Arabic language is structured and derivational language.
4. Arab peoples using their local dialect at home and within the community.
5. Using of diacritics instead of dedicated letters to represent short vowels.
6. Arabic letters can be changed depending on its position in the word.
7. The Arabic names do not use capital letters
8. There are no symbols used to identify sentence boundaries.
9. The conflict of using diacritics in Arabic text.

4.5 Arabic language ambiguity

An ambiguity concept in linguistics means one word has two meaning or more, or one pronunciation or more. Ambiguity in language is one of the important research area in computation linguistics. Many studies of the ambiguity were done, but one of the important studies done in Arabic language done by [77] which state that there are levels of the ambiguity (in linguistic domains; phonological, morphological, syntactic and semantic levels), another taxonomy of ambiguity mentioned by [78], those types of ambiguity are: syntactic, lexical, structural, anaphoric (repetition of a word or use of a linguistic unit, such as a pronoun, to refer to the same person or object as another unit), ellipses (omission of a word or words) and semantic, also [71] added and explained another type of ambiguity which is Constituent boundary ambiguity. Also [67] states one of the important features that Arabic language doesn't have, which is capitalization property; which specifically impacts and confuses the recognizable proof of the Arabic Named Entities. We will focus on morphological, syntactic and semantics levels.

4.5.1 Arabic language morphological ambiguity

There are two strategies to deal with a morphological level, which are stem-based morphologies, and root-based morphologies. One of the main problems in Arabic language computation is a morphological problems, because until now there is no good solution specially in POS taggers , syntactic parsers, and machine translation[77]. In the following part, we give some examples to explain some morphological ambiguity in the Arabic language. [77] explained the sources of morphological ambiguity which are:

1. Changing in the orthographic operation produces inflected form that can belongs to two stems or more .
2. When we double sounds in the same lemma, it pronounced but not written.
3. When we use diacritics, it can change the meaning of the same word.
4. Adding some prefixes and some suffixes can be homographic with each other.
5. Adding some prefixes and some suffixes to the word can produce an original homograph.
6. Uninflectional words (homographs) can give different meanings.

For more details return to [77].

4.5.2 Arabic language syntactical ambiguity

Syntactic ambiguity concerned with sentence structure, most studies in Arabic language ambiguity were done in morphological ambiguity, but there are some studies in syntactical ambiguity, one of these important studies that one made by [79] which explains the source of difficulties in Arabic parsing, these difficulties are:

1. Arabic sentences are complicated and long.
2. The sentence structure is complex because there are some factors affecting such as: grammatical relations, order of words and phrases.

One of the main problems is sentences order, [80] clarified that the basic order is VSO (verb, subject, object), and VOS, SVO, and VO orders are also possible.

4.5.3 Arabic language semantic ambiguity

[78] explained that "Semantic ambiguities coming from polysemous lemmas", this phenomenon is called [81] lexical semantic ambiguity. [82] explains many models were done for semantics Ambiguity and those models made mapping word matching one word and this wrong. Also [83] explained that ambiguous words have multiple entries within a lexical network. Also [84] explained the role that semantics plays in Arabic language processing in the relationships between words, he stated some of the important relationships between words which must be considered, these relationships are: synonym المترادفات, antonym المضاد, hypernym , hyponym الفرع, holonym تحتوي على , meronym توجد في and the type of word (POS).

4.5.4 Some of the proposed solutions to resolve the problem of linguistic ambiguity in the Arabic language

There are some proposed solutions for language ambiguity in general and particular in Arabic language, there are general solutions explained by [78] those solutions are:

1. Lexical ambiguity: it can be solved by the syntactic analyzer.
2. Syntactic ambiguity: it can be solved by the semantic analyzer.
3. Structural ambiguity: it can be solved by the semantic analyzer.
4. Anaphoric ambiguity: it can be solved by syntactic analysis within a sentence.
5. Semantic ambiguity: it can only be solved at the semantic level.

Also [85] mentioned the morphological problem can be solved by rule-based or stem-based and finite state morphology. Also [86] mentined that a lexicon has two features : syntactic features to resolve syntactic ambiguity and lexical features to resolve lexical ambiguity. Another solution for semantics ambiguity is presented by [78] which is used graph to represent each word and to retrieved from semantic lexicon. A new methodology designed by [65] which ontology for concepts. Another solution used in question answering system is mentioned by [87] which states that some systems exploit general taxonomy for question (who, where, when, etc.) for semantic classes.

4.5.5 Arabic language tools

Arabic language is one of languages that rich in traditional resources[84], but there is a lack of Arabic digital language resources[88], and most of these tools designed for Roman language and suitable for Arabic language[89], and the computational community take advantage of building these resources[66]. These resources are intelligent and non-intelligent [90]. Also these resources are available internal or external[91]. Many Arabic language tools were proposed, we will explain some of these tools (Most of these tools mentioned above in pages 11-12) in the following part briefly:

1. **Stemming**: is the process extract affixes from the word, many stemmer were developed for Arabic language such as: Khoja stemmer for root extraction and light stemmer for removing prefix and suffixes[92].
<https://sourceforge.net/projects/arabicstemmer/files/latest/download>
2. **Corpus**: It is a collection of examples in real world texts, one of the important corpuses is KSUCCA[93] which is designed by King Saud University and KACST [94] which designed for King Abdulaziz City for Science and Technology.
<http://arabiccorpus.byu.edu/>
3. **Arabic Part of Speech Tagging**: Arabic language word carries clitics and inflectional[95] for this we use Part of Speech Tagging (POS) as in the Arabic word وسياكلونها (wasayakolonaha) the root is أكل akl and it consists of following : a conjunction, a future particle, an inflectional prefix the verb stem, an inflectional suffix, and a pronominal object.
<http://nlp.stanford.edu/software/stanford-postagger-full-2015-12-09.zip>
4. **Arabic word net**: We need to use a tools to to deal with Arabic word disambiguation, which is defined as "the task of automatically assigning a meaning to an ambiguous word in a given context" [96]. More than 40 wordnets (WNs) were building[97]. For the Arabic language Arabic word net is built based on Princeton WordNet for English[98]. One of the important application in this area is Word net Browser [99] is a Java stand alone application used to searching for concepts and to modify data from lexicographers.

5. **Parser**: parsing is the process of analyzing sentences or string of symbols by using grammar rules. It consist of three parts: lexicon, morphology, and syntax[100]. Many parser were vailable, but one of these most used parser is Stanford parser , which is available in:
<http://nlp.stanford.edu:8080/parser/>.
6. **Treebank**: it is a text f corpus that annonated to explain syntactic and semantic structure. One of the important Treebank is Penn Treebank[101], which is available at: <https://www.cis.upenn.edu/~treebank/>.
7. **Spell checker**: it is the process of checking the words characters are right or may be wrong, many tools were built such as:
https://www.spellchecker.net/arabic_spell_checker.html.

4.6 Holy Quran

4.6.1 Introduction

Holy Quran is the word of Allah, carried by the angel Gabriel to prophrt Mohammed 1355-1378 years ago[90], it is written in (C 600. CE) in classic Arabic[102]. It is the central religious text of islam [103]. [68] mentioned that it is the source of religious legislation for Muslims.Holy Quran contains many knowledge which ranges from Humanities such as history, legal and law such as moral and science[63]. It cosists of 114 surah (chapters) and 2636 Ayah (verses)[65] and varies in length [104], it is made up of 77,430 words [105], it is divided into 114 chapters which and each surah has its name [106]. Muslims peope take their inspirations and beliefs from two sources: the Quran and Sunnah[107]. Also [108] states that all Muslims should recite Holy Quran in Arabic language regardless of their native language.[67]

4.6.2 The role played by the Holy Quran to save Arabic language

The main language of the Holy Quran is the Arabic language. Holy Quran is the miracle of the prophet Mohamed (peace be upon him), which challenges the Arabs to produce the like of the Quran, or Surah or one verse, and they could not so far. [109] explains that Holy Quran save Arabic language and stated that Arabic

language is a live language because it's resistant to the causes of extinction and to it coincided with the evolution of semantics add to their wording.

An other one mentioned that[110]: there were no languages immortalized by its book, but only the Quran, which is immortalized the Arabic language.

4.6.3 Holy Quran sciences

Many sciences of the Holy Quran mentioned by [111] Those are: recitations, interpretations, scientific Mutoon, memorization, etc. Many books written in Holy Quran sciences, and each one categorized it into many chapters depending on many characteristics, these categorizations in general are the same, but there are some omission and additions in each category by the authors. Here are some references and categories, and their sciences:

1. 50 types of sciences. مواقع العلوم من مواقع النجوم للبلقيني.
2. 47 types of sciences. البرهان في علوم القرآن للزركشي.
3. 80 types of sciences. الإتيان في علوم القرآن للسيوطي.
4. 154 types of sciences. الزيادة والإحسان علوم القرآن لابن عقيلة.

These categories are classified depending on many characteristics. We can explain some of these types depending on category 1 [112] such as:

1. The reasons for the descent of the Quran and its times and its Proceedings such as: Mekki المكي which is a descence in Makkah, Madani المدني which is a descent in Madinah, on travel, on urban, at night, at midday, the first decent and the last decent, etc.
2. The Sanad السند: it is the science of the way leading to the Matn(Is what ended up talking to him said or done or report to the Messenger of Allah in Islam Mohammed bin Abdullah). There are many categories such as: Frequent المتواتر, the ones الأحاد, abnormal الشاذ, narrators الرواة.
3. Performance in reading such as: stopping الوقف, starting الابتداء, slurring الإدغام, Hamza الهمزة.
4. The Words: Strangely الغريب, inflected المعرب, common المشترك, synonyms المترادفات, metaphor المجاز, simile التشبيه.
5. Meanings related to judgments.
6. Meanings related to words.

4.6.4 Holy Quran applications

Last decades holy Quran takes attention from scholars (computation, linguistics and Islamic scholars), but until now there has been a lack in scientific Islamic and holy Quran application, in the following parts we will explain some of these applications briefly.

The first application is project Ayat (آيات) ⁴, which is designed in for King Saud University, designed for PC, mobile and social media Facebook and twitter. Its resources were grouped from multiple sources. It has the following features: search, Quran interpretation, listening to the holy Quran. Its main problem is the meaning of words and verses are not revised.

The second application is Quran flash⁵ which is designed by volunteers and then commercial. It has characteristics of using many types of reading Quran (Qira'at), many types of Mashaf (Madinah and Makkah), display the types of chapters (Mekki or Medani graphically) and one of the important characteristics is: zoom.

4.7 Chapter summarization

This chapter gives general introduction to Arabic language and its characteristics, Arabic NLP, Arabic language challenges, Arabic language ambiguity and some proposed solutions, Arabic language tools. Also Holy Quran investigates: its roles to save Arabic language, its science and applications.

⁴ http://quran.ksu.edu.sa/index.php#aya=1_1&m=hafs&qaree=husary&trans=ar_mu.

⁵ <http://www.quranflash.com/books/Medina1/?ar&aff#>

CHAPTER V

5 Chapter five: The experimental design and results

Our main goal of this thesis is to build question answering system for holy Quran, that can help Muslims to find answers from holy Quran, and to enhance this system to find good results to help researchers in this field.

In this part, explains tools used, methodology, building corpus, building six prototypes each one have some characteristics different from others, and discussing results.

5.1 Tools and methodology

Many tools and applications used in this project. These tools are

5.1.1 Java programming language

Java is one of the important programming languages and computing platform for many applications. It's released by Sun Microsystems in 1995. Many applications use Java specially web application high potential in the field of programming web, games, databases, and many other applications. It has many compilers and editors such as text pad, Eclipse platform, and NetBeans platform. In addition, Java has the following characteristics:

- Secure.
- Fast.
- Reliable.
- Java works in many machines such as laptops, datacenters, game consoles, supercomputers, cell phones.

There are a lot of applications and websites that will not work, unless you have Java installed, and more are created every day. From laptops to datacenters, game consoles to scientific supercomputers, cell phones to the Internet, Java is everywhere!

We use Java program to connect with Microsoft access to search and retrieve the required information from (verses from chapter or the relevant files which contains the required information) the database or the files from the directory to the user interface.

5.1.2 Microsoft Excel

Microsoft Excel is an electronic spreadsheet program. It has replaced Lotus, which is the industry standard spreadsheet earlier. Excel is part of Microsoft Office. It used for storing, organizing, and manipulating data. It consists of one or many sheets. Each one is a table that contains cells, and each cell represents the intersection of a row and a column. We used it for storing, organizing and manipulating data for building our corpus. In addition, we use it to build some charts and some important calculations.

5.1.3 Microsoft Access

It is a relation database engine produced by Microsoft with graphical user interface and software development tools. It is a part of Microsoft office tools. Access supported many programming languages such as Java and visual Basic. It used for storing and manipulating many types of data.

In this project, it used to store our questions and their results (chapters and verses) , then the user can access the right result for the query after many processing needed by a Java program.

5.1.4 Lucene search engine

Apache Lucene is a text search engine with high performance, full search featured library written in Java, it is suitable for nearly any application that requires full-text search, especially cross-platform which is an open source project available for free . It can be enhanced and modified depending on user requirements to be more efficient. The project used it to stem words and designing patterns to enhance the results of search.

5.1.5 Diagram designer

Diagram Designer is a soft ware that helps in creating flowcharts, diagrams. It has many features of easy editing, object manipulation free and customizable.

5.2 Building data set for holy Quran questions and answers

Our first task was extracting an example set of questions and answers from the Holy Quran (Surat Al-Fatiha and Al-Baqarah Chapters) to be the core of a

question and the answers corpus to be used as a gold standard corpus. Frist part explains the main and general steps of designing the corpus and then more details

5.2.1 Procedures of collecting data:

This corpus collected and prepared from many resources, many steps were done, these processes are:

1. Collecting questions from real users.
2. Collecting questions from the internet.
3. Combining the two groups of questions in one resource Microsoft excel.
4. Collecting answers from AlFatihah'h and AlBagarah verses from *مصحف المدينة النبوية للنشر الحاسوبي*.
5. Combining all these data set in many sources: first one: each verse in a text file, second one: Access database containing the verses, answers, some other details, third one is csv file.

5.2.2 Types of data:

The data type is text collected from the internet and written text from real users.

5.2.3 Clean the data:

The data were collected, have many problems ,which are: writing Holy Quran verses use diacritics to clarify how to read this verse, Holy Quran uses special characters, also data collected from internet uses some time English characters in Arabic texts. To solve these problems, many processes were done which are:

Removing diacritics and some symbols were removed using Microsoft Access tool Find and replace which is found in editing part: some characters replaced with others such as ^أ with ^ا, some English characters were removed from the text such as : u, some diacritics were removed by using find the diacritics and replace it with nothing.

5.2.4 Details of building corpus:

This work was done by reading the holy Quran and devising suitable questions relating to each verse and its answer, which is the verse number in this chapter or any other related verses number(s) in other chapter(s). These questions and answers written in Excel spreadsheet, see table [1]below.

Table 5—1: Quran QA spreadsheet header

الرقم NO	السؤال Question	السورة hapter	الآية Verse	رقم السورة Chapter No
-------------	--------------------	------------------	----------------	-----------------------------

Our example dataset contains 215 questions and more answers because several verses can be answers to one question. Also the last column is for comments to explain some verses when it needs more explanation to be clear. These questions were then validated and revised in Gabrah College by Islamic and Arabic scholars. Table [2] has some examples from the validated spread sheet.

Table 5—2: Quran QA sample after expert validation

الرقم	السؤال	السورة	الآية
1	لمن الحمد؟	1	2
2	ما هي صفات الله التي ذكرت في سورة الفاتحة؟	1	2-3-4
99	أين يؤتى النساء؟	2	222
104	هل الرجال والنساء على درجة واحدة ولم؟	2	228
210	هل يحاسبنا الله على ما نبديه أو نخفيه في أنفسنا؟	2	284

A second example Quran QA dataset containing 47 questions was extracted from an Islamic web site. In this source there are some problems in presenting data, as it used some English letters like r and U. We extracted the questions and related answers from the website as illustrated in Table [3].

Table 5—3: Quran QA sample from WWW

الرقم	السؤال	السورة	الآية
1	ما هي السبع المثاني؟	1	1-2-3-4-5-6-7
7	من هن الأزواج المطهرة؟	2	25
16	حتى ماذا كان شرط قوم موسى يؤمنوا في أن الكلام الذي يسمعه هو كلام الله؟	2	55
21	لأولاده ماذا كانت وصية إبراهيم لأولاده؟ وكذلك وصية يعقوب	2	132

Finally we combined the two sources in the third spreadsheet containing 263 Quran questions and their answers. This spreadsheet was sorted by verse numbers. It was easy to check there is no duplication here because we only covered two chapters of the Quran; but if we continue there may be duplications and so we may need another column as a cross-check key to show the chapter number and the verse or verses. Each answer that includes more than one verse was marked in the end of the rows. With the combined spreadsheet the data was processed as follows:
To differentiate the two different sources we changed the font type and size as in Table [4].

Table 5—4: Differentiate between sources

لمن الحمد؟	1	2
ما هو الكتاب الوحيد الذي لا يوجد أي ريب أو شك فيه؟	2	2

If the question and its answers are found in both the two sources: colored yellow as in Table [5].

Table 5—5: repeated answers and questions from many resources

15	كم عدد السماوات؟	2	29	مكرر
16	ما عدد السموات؟	2	29	مكرر

If there are two questions repeated and followed by another two repeated questions we colored first yellow and second blue as in Table [6].

Table 5—6: double repeated question from different sources

15	كم عدد السماوات؟	2	29	مكرر
16	ما عدد السموات ؟	2	29	مكرر
17	ماذا علم الله نبيه آدم؟	2	31	مكرر
18	ماذا تعلم آدم عليه السلام من الله جل جلاله وكان هذا العلم ليس عند الملائكة ؟	2	31	مكرر

If there are three questions repeated from the two sources but the second and the third interleaved in spreadsheet sorting we colored them as yellow and blue and red as shown in Table [7].

Table 57—: Three repeated question from different resources

34	2	من هو الذي لم يسجد آدم ؟	20
34	2	من الذي رفض أمر الله تعالى بالسجود لآدم عليه السلام؟	21
35	2	أين أمر الله آدم وزوجه أن يسكنا؟	22
35	2	أمر الله تعالى آدم وزوجه أن لا يقربا شيئاً ما هو ؟	23
35	2	عندما خلق الله تعالى آدم عليه السلام وحواء ماذا كان سكنهما ؟	24
35	2	ما الشيء الذي منعه الله تعالى عن آدم وحواء في الجنة ؟	25

Finally we add a new column for verse English translation by Abdullah Yusuf Ali, who "was a British-Indian Islamic scholar who translated the Qur'an into English. His translation of the Qur'an is one of the most widely known and used in the English-speaking world. He was also one of the trustees of the East London Mosque." Our final version of the Quran QA corpus is illustrated in Table [8].

- Stemmer: it is used to produce related words to the key word by adding some pattern to the word to expand the query.

5.3.2 The process steps in the model

The following parts explain process steps for this model.

- Input user question: to read a question from user in natural language.
- The question checked in the database question to check if this question is entered before?
 - I. If the answer is yes, then the system retrieves its answers from answers database and display it.
 - II. If the answer is no, then the system store the question in the question database and index it. Also the system must do the following processing:
 - 1) Tokenizer: to split the question into words to be key words for searching.
 - 2) Question classifier: to classify question in many categories depending on many characteristics
 - 3) Stemmer: is a technique used to remove additional characters from inflected or derived word to its original word stem (root).
 - 4) Replacer: is the process of replacing some character with another, such as: $\dot{\text{ا}}$ with ا in Arabic language.
 - 5) Remover: to remove some characters, diacritics, symbols, and signs.
 - III. After these processes, the system processes the query words to expand it depending on many resources such as: dictionary, corpus, thesaurus, and new patterns.
 - IV. The system displays the words of last query to check by the user to select the suitable words for the query by removing the unwanted words.

- V. The system uses these words to search for documents internally in a corpus (off line) or externally in web resources (on line).
- VI. The system extracted answers from documents and ranked it.
- VII. Answers displayed to check it if valid index the answer and store it in database answers. Else terminate.

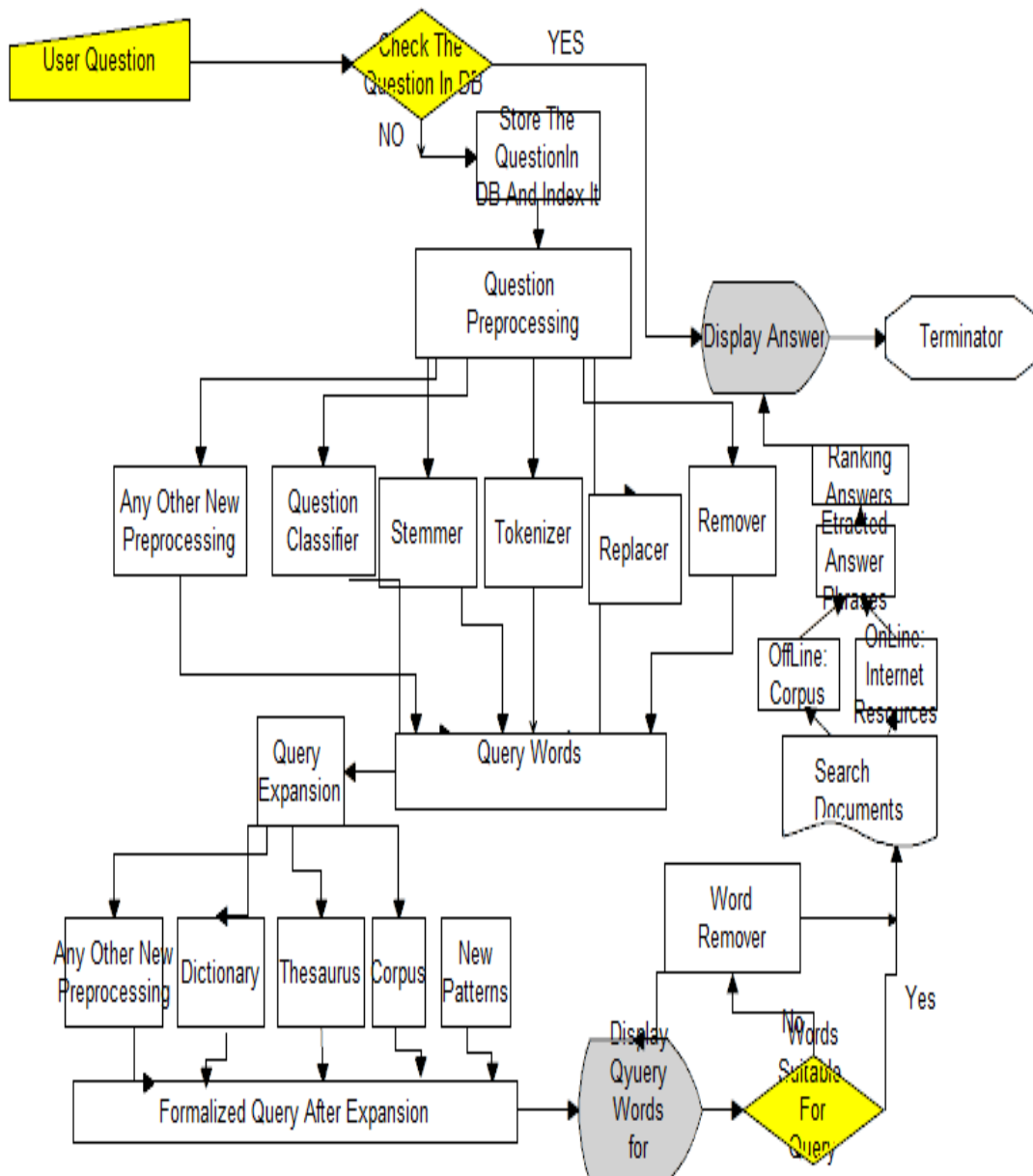


Figure 5—1: The proposed framework for Arabic question answering system.

5.3.3 Framework characteristics

Many characteristics distinguish this framework such as:

1. This framework has the three main parts of the question answering system query processing, document retrieval, and answer extractions.
2. Also, it has the ability of interaction with users to make the system easier.
3. Other characteristics this framework can be used for information retrieval.
4. It can be extended by adding other features and NLP methods.
5. The framework can be applied in many areas such as: a question answering system for Holy Quran, a question answering system for communication systems for customers help.
6. We can use the framework to build many corpuses for each area by adding any new data.

5.3.4 Drawbacks of framework

The framework has the problem of heavy load if we apply all these techniques, also in query expansion, there is also another load.

5.4 Building prototypes for question answering system

In this part six programs were built(prototypes), then making comparison between these prototypes and discussing these results. Also enhancement done for these programs and make another comparison and discuss these results.

5.4.1 The First prototype

We built our prototype using Java standard edition version 8, the following figure[5-2] explains the main steps .

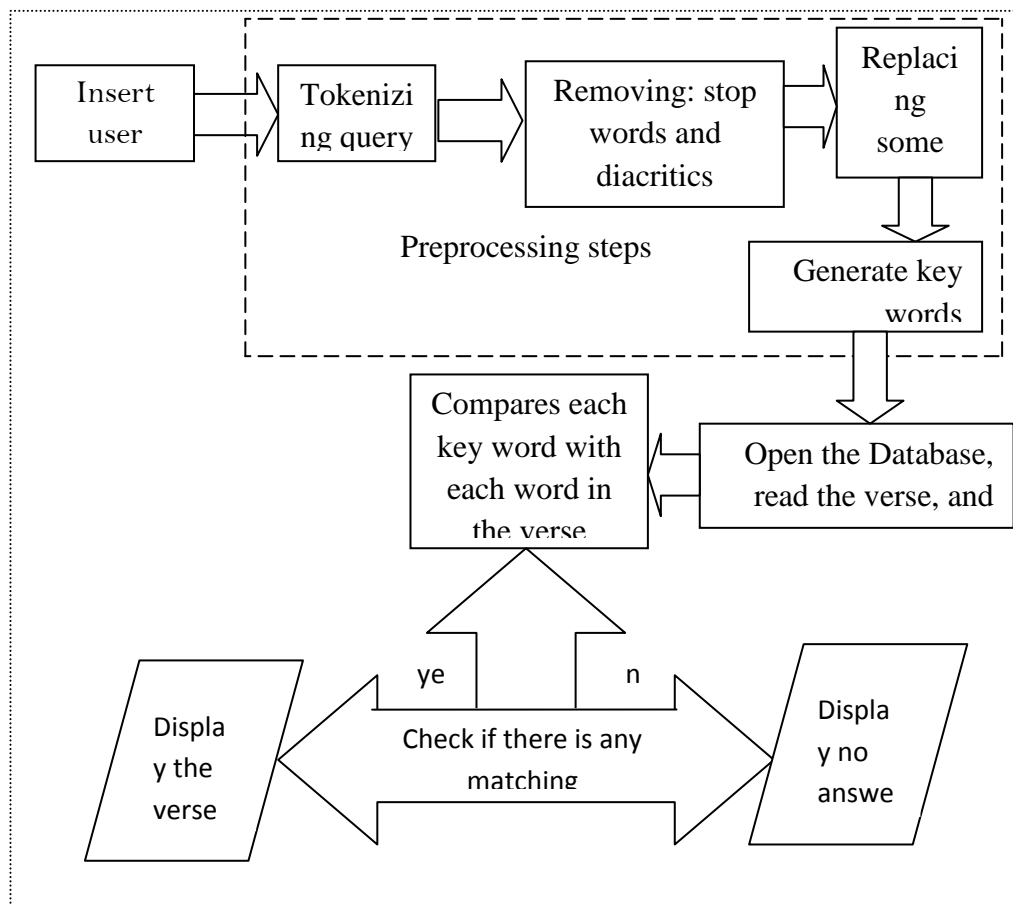


Figure 52—: main processes of prototype 1

The system involves the following steps:

1. The input string (user query) is tokenized into a string for each word.
2. Use tokenized strings as keywords for search.
3. The tokenized words are stored in an array of string
4. Then the tokenized words are compared with the database (after making the connection and storing their fields in variables), each one with the all words in the record; if any word matches then display the record.

The system works as follows:

1. Accepting a question as in figure[5-3].

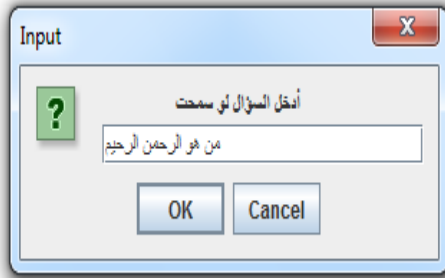


Figure 5—3: Entering the question form

Then the above processing steps must be done as follows:

2. The string is tokenized into words.
3. Using the tokenized keywords for search.
4. Then the tokenized words are compared with the database, each one with the all words in each record (verse). If any word matches then display the record. This may lead to unnecessary results as in the following figure [5-4] (only part of result).
5. Also at this stage if we use Modern Standard Arabic in our question which is the standard language in use today we may not find direct matches to Quran verses (we can use the English translation and the question as an evidence) because verses use diacritics. If we search about: *الحي القيوم* there is no result as in Figure [5-5] and [5-6].

اسم السورة البقرة وَمَلَايِكْتِهِ وَكُتُبِهِ وَرُسُلِهِ لَا نُفَرِّقُ بَيْنَ أَحَدٍ مِنْ رُسُلِهِ وَقَالُوا سَمِعْنَا وَأَطَعْنَا غُفْرَانَكَ رَبَّنَا وَإِلَيْكَ الْمَصِيرُ [The, Messenger, believeth, in, what, hath, been, revealed, to, him, from, h	رقم الآية 285
اسم السورة البقرة يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَأْخُذْكُمْ أَثْمَانُ وَالْأَسْخَابُ عَلَيْنَا أُولَئِكَ يَلْمُوكَ وَاللَّهُ خَبِيرٌ بِمَا تَعْمَلُونَ [On, no, soul, doth, Allah, Place, a, burden, greater, than, it, can, bear., It,	رقم الآية 286
عدد الإجابات = 151 BUILD SUCCESSFUL (total time: 15 seconds)	

Figure 5—4: results without removing diacritics and stop words

اسم السورة البقرة	رقم الآية 160
إِلَّا الَّذِينَ تَابُوا وَأَصْلَحُوا وَبَيَّنُّوا فَأُولَئِكَ أَتُوبُ عَلَيْهِمْ وَأَنَا التَّوَّابُ الرَّحِيمُ	
[Except, those, who, repent, and, make, amends, and, openly, declare, (t	
اسم السورة البقرة	رقم الآية 163
وَالَهُكُمْ إِلَهٌ وَاحِدٌ لَا إِلَهَ إِلَّا هُوَ الرَّحْمَنُ الرَّحِيمُ	
[And, your, Allah, is, One, Allah., There, is, no, god, but, He., Most, Gra	
عدد الإجابات = 5	
BUILD SUCCESSFUL (total time: 30 seconds)	

Figure 5—8: The answer after removing stop words and diacritics

7. We notice that the number of results becomes 5 instead of 151. Also if we removes punctuations from verses as well we find results instead of no results as in Figure [5-9] and [5-10].

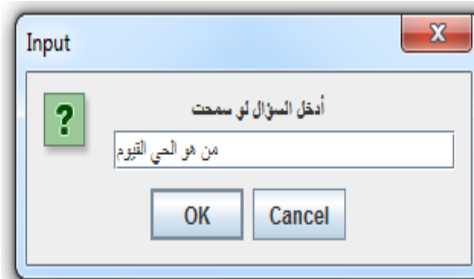


Figure 5—9: the last question

اسم السورة البقرة	رقم الآية 255
تِ وَمَا فِي الْأَرْضِ مَنْ ذَا الَّذِي يَشْفَعُ عِنْدَهُ إِلَّا بِإِذْنِهِ يَعْلَمُ مَا بَيْنَ أَيْدِيهِمْ وَمَا خَلْفَهُمْ وَلَا يُحِيطُونَ بِشَيْءٍ	
[Allah., There, is, no, god, but, He,-the, Living,, the, Self-subsisting,, Etern	
عدد الإجابات = 1	
BUILD SUCCESSFUL (total time: 15 seconds)	

Figure 5—10: the result after removing diacritics and stop words and punctuation

A number of observations can be summarized as follows:

- Tokenization is a very important stage because we need to make comparisons between each word in the question and each word in the verses.
- Removing stop words and diacritics increases the efficacy and efficiency of the system.
- The link between the answers and any other fields in the database table can decrease answers; for this we remove the question as evidence in search, but we can add another field to increase the efficiency such as أسباب النزول.
- The system needs a lot of memory because it uses a lot of strings and arrays of strings; the system stores the entire Quran QA database in memory.
- Finally, we summarize the results in table 9, which shows the increase of the matching answers after removing diacritics (little) and stop words (more).

Table 5—9: Table of first comparisons

question	Number of results recalls and its percentages			
	No removing Stop words and diacritics	removing Stopwords only	removing diacritics	removing Stop words and diacritics
من هو الرحمن الرحيم Who is the Most Most Gracious and . Merciful	0 = 0%	0 = 0%	151 =3.3%	5 =100%
من هم بني اسرائيل Who are the children of Israel	0 = 0%	0 = 0%	163=4.2%	7=85.7%
ما هو الصيام What is fasting	0 = 0%	0 = 0%	70=2.85	2=100%

من هو الحي القيوم Who is living and subsisting	0 = 0%	0 = 0%	147=0.68 %	1=100%
--	--------	--------	---------------	--------

5.4.2 The second prototype

The second prototype depends on indexing, which used by Lucene search engine. The system search through our corpus, which contains the following files (Holy Quran: Al-Baqarah and Al-Fatihah chapters, verses of each chapter).

Searches using Lucene search engine packets, which is used reverse indexing which indexes words in files and give each word index to tag the words and files that contain these words , table 10 give examples of indexing process in the second prototype:

Table 5—10: Example of indexing some words by indexing system

Word	Documents contains the words
الرحمن	1-1.txt, 1-3.txt, 2-163.txt
الرحيم	1-1.txt, 1-3.txt, 2-37.txt, 2-128.txt, 2-160.txt, 2-163.txt, 2-182.txt 2-192.txt, 2-199.txt, 2-226.txt
الحج	2-150.txt, 2-189.txt, 2-196.txt, 2-197.txt
الشيطان	2-36.txt, 2-168.txt, 2-208.txt, 2-268.txt, 2-275.txt

Also , figure[5-11] shows part of the indexing result of verses files.

```

adding C:\Users\adany\Desktop\project\alfateha\2-91.txt
adding C:\Users\adany\Desktop\project\alfateha\2-92.txt
adding C:\Users\adany\Desktop\project\alfateha\2-93.txt
adding C:\Users\adany\Desktop\project\alfateha\2-94.txt
adding C:\Users\adany\Desktop\project\alfateha\2-95.txt
adding C:\Users\adany\Desktop\project\alfateha\2-96.txt
adding C:\Users\adany\Desktop\project\alfateha\2-97.txt
adding C:\Users\adany\Desktop\project\alfateha\2-98.txt
adding C:\Users\adany\Desktop\project\alfateha\2-99.txt
3627 total milliseconds
BUILD SUCCESSFUL (total time: 8 seconds)

```

Figure 5—11: The indexing of holy Quran verses (Al-Baqarah and Alfatihah)

The following figure displays the second prototype processes as follow

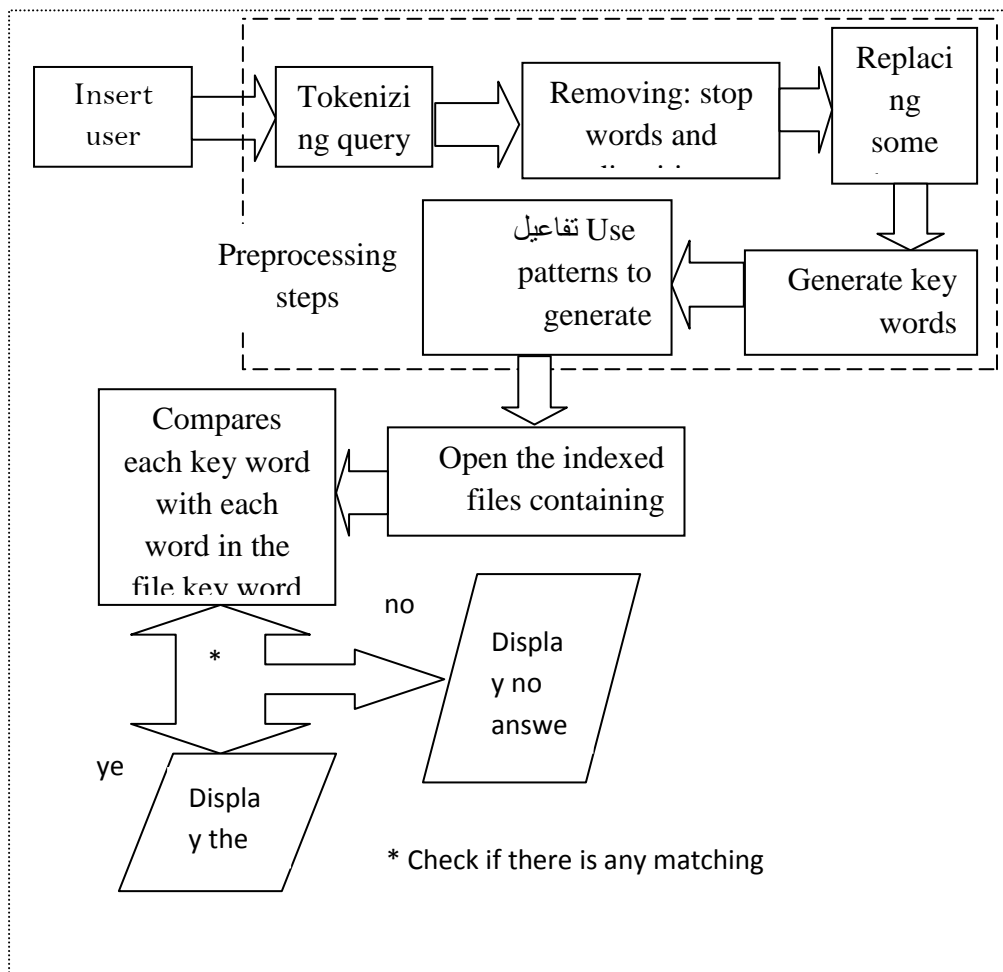


Figure 5—12: second and third prototypes processes

Then the system applies the following steps:

The user inserts its query through the search program as figure [5-13].

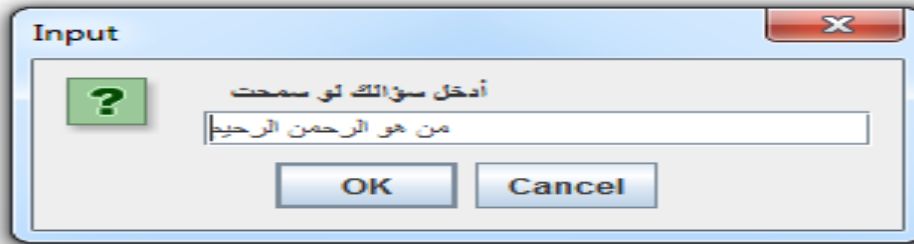


Figure 5—13: Inserting query in second prototype

Then the search program makes many processes for the query, these processes are:

1. Removing stopping words such as (من، هو، كيف , etc.).
2. Stemming the original words that the query uses to search.
3. Displaying these words after stemming, as in figure [5-14]:

السؤال بعد التعديل هو : رحمن رحيم
14 total matching documents
1. C:\Users\adany\Desktop\project\alfateha\1-3.txt
□الرَّحْمَنُ الرَّحِيمُ (3)
2. C:\Users\adany\Desktop\project\alfateha\1-1.txt
□بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ (1)
3. C:\Users\adany\Desktop\project\alfateha\2-163.txt
□وَإِلَهُكُمْ إِلَهٌ وَاحِدٌ لَا إِلَهَ إِلَّا هُوَ الرَّحْمَنُ الرَّحِيمُ (163)

Figure 5—14: The main words for the query after stemming some letters

4. Then the system search index to find these words, if it found then the system reads the file that corresponding to these words and display its contents and as in figure[5-15].

8. C:\Users\adany\Desktop\project\alfateha\2-37.txt

□

فَتَلَقَىٰ آدَمُ مِنْ رَبِّهِ كَلِمَاتٍ فَتَابَ عَلَيْهِ إِنَّهُ هُوَ التَّوَّابُ الرَّحِيمُ (37)

9. C:\Users\adany\Desktop\project\alfateha\2-128.txt

رَبَّنَا وَاجْعَلْنَا مُسْلِمِينَ لَكَ وَمِنْ ذُرِّيَّتِنَا أُمَّةً مُسْلِمَةً لَكَ وَأَرِنَا مَنَاسِكَنَا وَتُبْ عَلَيْنَا إِنَّكَ أَنْتَ التَّوَّابُ الرَّحِيمُ (128)

10. C:\Users\adany\Desktop\project\alfateha\2-182.txt

□ فَمَنْ خَافَ مِنْ مَوْصٍ جُنْفًا أَوْ إِثْمًا فَأَصْلَحَ بَيْنَهُمْ فَلَا إِثْمَ عَلَيْهِ إِنَّ اللَّهَ غَفُورٌ رَحِيمٌ (182)

BUILD SUCCESSFUL (total time: 2 minutes 3 seconds)

Figure 5—15: The result of the query من هو الرحمن الرحيم

5.4.3 The third prototype

The third one builds by adding new patterns for irregular plural (broken plural (جموع التكسير) depending on Lucene search engine (second prototype) applying in Quran verses corpus, in this part we use these patterns (فاعيل (فاعيل، فياعيل) instead of one weigh in prototype 2 which is تفاعيل to increase the precision of results, because the system return each broken plural to its singular and use all the words that produced as a new words for searching. We find only one word in AlBagarah and Elfatihah chapters, which is تصريف the plural word is then pattern is تفاعيل for more details return to figure[5-16]. When we use the word تصريف which is singular and appeared in AlBagarah chapter verses 164. We use the query: هل للرياح تصريف in all prototypes we find the following results as in figures [5-16]:

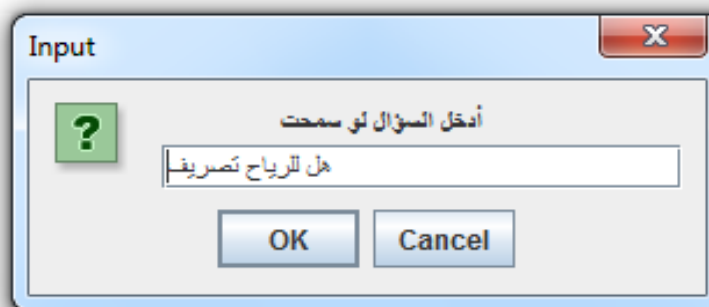


Figure 5—16: entering the query "هل للرياح تصريف"

1. The first prototype result

Your Question is :

هل للرياح تصريف

رقم الآية 164
اسم السورة البقرة
مِنَ السَّمَاءِ مِنْ مَّاءٍ فَأَخْبَا بِهِ الْأَرْضَ بَعْدَ مَوْتِهَا وَبَثَّ فِيهَا مِنْ كُلِّ دَابَّةٍ وَتَصْرِيفِ الرِّيَّاحِ وَالسَّحَابِ الْمُسَخَّرِ بَيْنَ السَّمَاءِ وَالْأَرْضِ لآيَاتٍ لِقَوْمٍ يَعْلَمُونَ
[Behold!, in, the, creation, of, the, heavens, and, the, earth,;, in, the, alternation, of, the, night, and, the, day,;, in,

عدد الإجابات = 1

BUILD SUCCESSFUL (total time: 22 seconds)

Figure 5—17: The first prototype result for the query "هل للرياح تصريف"

2. The second prototype result:

السؤال بعد التعديل هو : هل رباح تصريف

2 total matching documents

بَعْدَ مَوْتِهَا وَبَثَّ فِيهَا مِنْ كُلِّ دَابَّةٍ وَتَصْرِيفِ الرِّيَّاحِ وَالسَّحَابِ الْمُسَخَّرِ بَيْنَ السَّمَاءِ وَالْأَرْضِ لآيَاتٍ لِقَوْمٍ يَعْلَمُونَ (164)

□ هَلْ يَنْظُرُونَ إِلَّا أَنْ يَأْتِيَهُمُ اللَّهُ فِي ظُلَلٍ مِنَ الْعَمَامِ وَالْمَلَائِكَةُ وَقُضِيَ الْأَمْرُ وَإِلَى اللَّهِ تُرْجَعُ الْأُمُورُ (210)

BUILD SUCCESSFUL (total time: 31 seconds)

Figure 5—18: The second prototype result for the query "هل للرياح تصريف"

There are wrong results because Lucene software have less stop words but, for this we need to enhance by adding some stop and common words that affects in search and give wrong results such as the Arabic word هل.

3. The third prototype result: when we use the plural تصاريف in the query we find the following results:

4. The first prototype result figure [5-19].

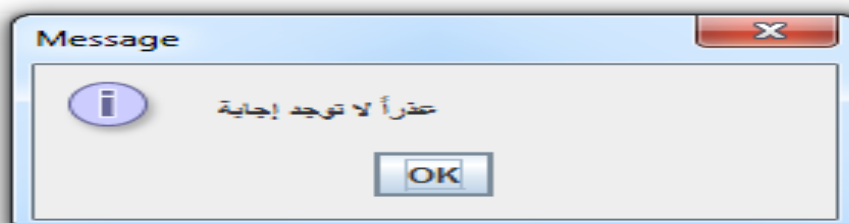


Figure 5—19: The result of question for the word تصاريف

5. The second prototype result

السؤال بعد التعديل هو : هل رياح تصاريف

2 total matching documents

□ هَلْ يَنْظُرُونَ إِلَّا أَنْ يَأْتِيَهُمُ اللَّهُ فِي ظُلَلٍ مِنَ الْغَمَامِ وَالْمَلَائِكَةُ وَقُضِيَ الْأَمْرُ وَإِلَى اللَّهِ تُرْجَعُ الْأُمُورُ (210)

وَمَا أَنْزَلَ اللَّهُ مِنَ السَّمَاءِ مِنْ مَاءٍ فَأَخْيَا بِهِ الْأَرْضَ بَعْدَ مَوْتِهَا وَبَثَّ فِيهَا مِنْ كُلِّ دَابَّةٍ وَتَصْرِيفِ الرِّيَاحِ وَالسَّحَابِ الْمُسْتَخَرِّ بَيْنَ السَّمَاءِ وَالْأَرْضِ لَايَاتٍ لِقَوْمٍ يَعْلَمُونَ (164)

BUILD SUCCESSFUL (total time: 16 seconds)

Figure 5—20: The result of question for the word تصاريف

There is an enhancing in results, because when the user enter the query that has the word تصاريف is plural of تصريف the word is reformulating and give the singular تصريف and used it as a new words added to the query as a key word for this a new answer appeared instead of no answer in the first prototype.

6. The third prototype result :

When we enter the question the system gives you 2 screens the first for the query after stemming figure [5-21] and the second for the singular patterns as in figure[5-22].

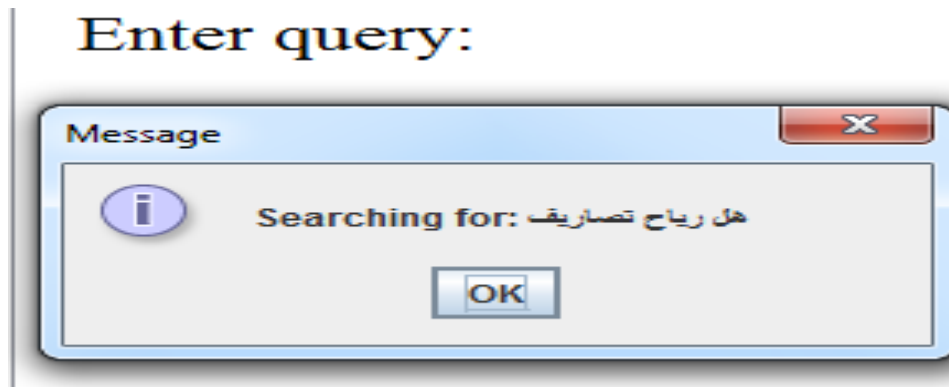


Figure 5—21: Displaying main query words

Enter query:

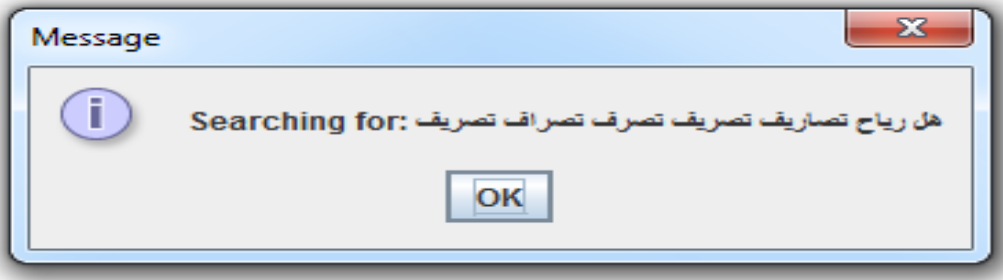


Figure 5—22: The derivatives morphological of the word تصاريق

السؤال بعد التعديل هو : هل رياح تصاريق

2 total matching documents

□ هل ينظرون إلا أن يأتيهم الله في ظلل من الغمام والملائكة وتضي الأمر وإلى الله ترجع الأمور (210)

وما أنزل الله من السماء من ماء فأحيا به الأرض بعد موتها ونبت فيها من كل ذبابة وتصريف الرياح والسحاب المنصر بين السماء والأرض لآيات لقوم يعقلون (164)

BUILD SUCCESSFUL (total time: 16 seconds)

Figure 5—23: The result of question for the word تصاريق and their derivatives morphological

5.4.4 Some important notices and observations for the three prototypes

From these results we notice that the first system give only one matching result, because it has many stopping words that decrease the precision of results, the first prototype removes the words: عليهم, هم from the query where the second prototype doesn't remove these words; for this there are many wrong results (37 from 38) and one matching. Secondly, the first prototypes uses database to store Quran verses where the second prototype uses indexing files and each one contains one verse. This makes the second prototype faster than the first because the systems scan only the files that contain the relevant words, where the first scans the entire database to find the relevant documents. Also the second [6] and the third prototypes [7] distinguish from the first prototypes by using some patterns that change the plural in broken plural to singular, also the third one have three patterns where the first one have only one pattern. These patterns help the system to check if the word in broken

plural in the weight of: فاعيل or تفاعيل or فياعيل then it returns to its singular. Unfortunately, in the two chapters were used as corpus AlFatihah and AlBagarah chapters we find only one verse matching these rules as in the above figures.

5.4.5 Experiments and results

In this part we discuss our experiments, which it judged By Scholars form Jabrah college. We use 30 question from corpus applied in the three prototypes. The following tables and figure, explain the results.

Table 5—11: General table results for the three prototypes

Question Number	Prototype (1)		Prototype (2)		Prototype (3)	
	Right answer	Matching	Right answer	Matching	Right answer	Matching
2	1	1	1	10	1	10
12	0	3	0	10	0	10
13	0	0	0	9	0	9
16	1	1	1	160	1	160
22	1	1	1	165	1	165
31	1	8	1	14	1	14
32	1	29	1	31	1	31
76	1	4	1	12	1	12
91	0	0	1	163	1	163
110	1	7	1	13	1	13
111	1	31	1	36	1	36

119	1	4	1	18	1	18
121	1	8	1	176	1	176
132	1	6	1	163	1	163
137	1	2	1	164	1	164
165	1	1	0	162	0	162
168	0	7	1	17	1	17
179	1	16	1	162	1	162
187	1	1	1	11	1	11
188	0	1	1	2	1	2
199	0	4	0	10	0	10
201	0	0	0	165	0	165
214	0	0	1	185	1	185
226	1	5	1	6	1	6
238	1	9	1	160	1	160
239	1	6	1	7	1	7
241	0	4	0	18	0	18
251	1	13	1	11	1	11
275	1	13	1	162	1	162
281	1	8	1	48	1	48
Summatio n	21	193	24	2270	24	2270

We use table [5-12] to generate the following tables and charts depending on the following formula:

$$\text{Percentage} = \frac{\text{number of right or wrong answer}}{\text{Total number of answers}} \times 100\%$$

Total number of answers

Table 5—12: Percentages of right and wrong answers

	Prototype (1)		Prototype (2)		Prototype (3)	
	Number	Percentage	Number	Percentage	Number	Percentage
Total of question	30	100%	30	100%	30	100%
Right answers	21	70%	24	80%	24	80%
Wrong answers	9	30%	6	20%	6	20%

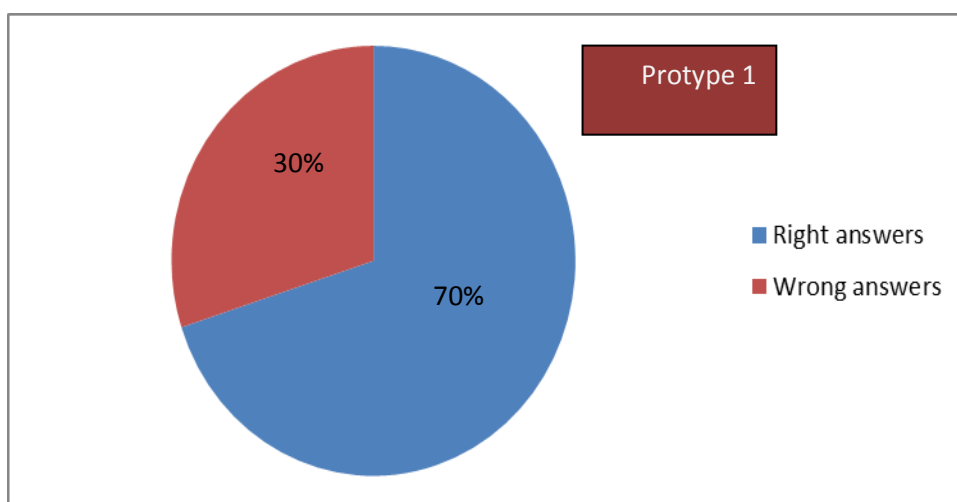


Figure 5—24: Displays percentages of right and wrong answer for prototype1

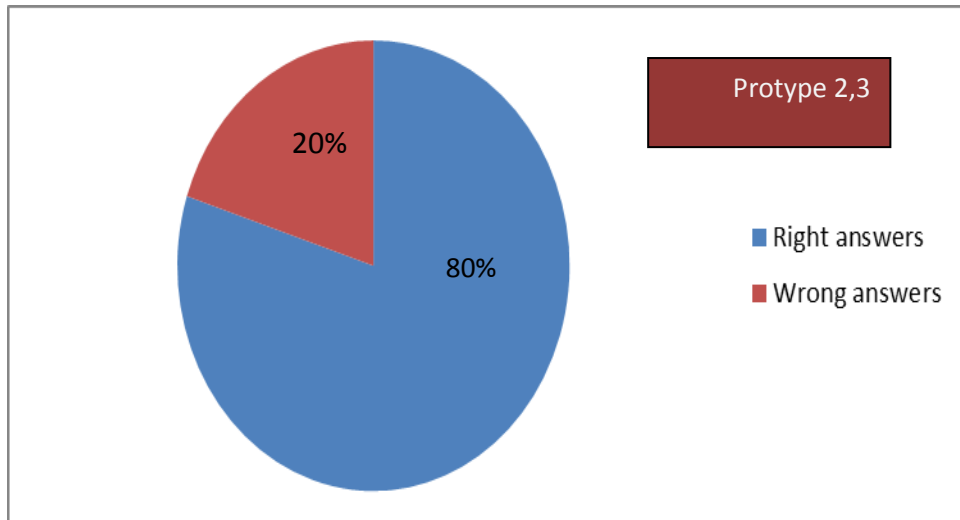


Figure 5—25: Displays percentages of right and wrong answers for prototype2 and 3

We notice that the prototypes 2 and 3 gives a more right answer because they used stemmer which increase the accuracy.

Table 5—13: Displays percentages of right and wrong answer in matching

	Prototype (1)		Prototype (2)		Prototype (3)	
	Number	Percentage	Number	Percentage	Number	Percentage
Right answers	21	10.88%	24	1%	24	1%
Wrong matching answers	193	89.11%	2270	99%	2270	99%

Also, we notice that in matching there an increasing in the precision results and vice versa the stemming decrease the precision of results.

5.4.6 The fourth prototype (Enhanced prototype1)

We built a new prototype depending on the framework proposed in figure[1-1] and prototype1, to add new features and enhance the prototype1. We use only the feature of checking the question if it answered or not, if the first choice, the question

answered directly without many linguistic processes, else applied the linguistics and search processes if there is an answer display it, else display no answer and save the question in a file. All the previous prototypes they don't have this feature (storing the question of the users), but this prototype has this feature. The following figure displays these details:

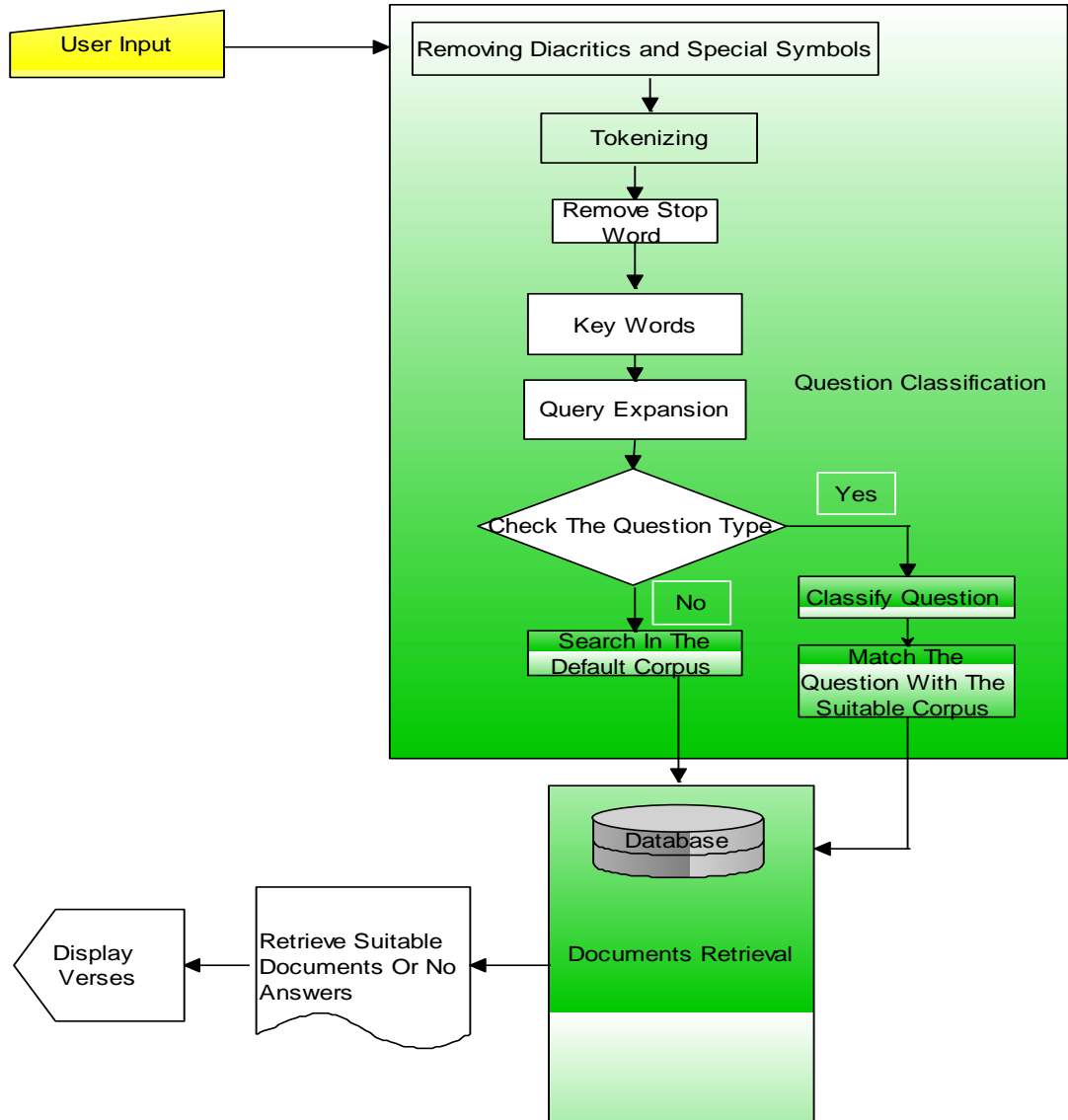


Figure 5—26: depicts prototypes 4 processes

The following steps discusses these processes in details.

1. Enter the question and then question stored in a string.

2. Storing the stop word in string (one line) instead of storing these stop words in many strings, each one in string (it takes about 60 line may be increase).
3. Check string is empty or not.
4. We use a function to remove stop words.
5. Removing diacritics such as : َ, ُ, ِ and replace other characters such as: í with \, ÿ with \, which they are stored in a string .
6. Tokenizing the string to words.
7. Check these words with question if there matching then display the answer directly; else
8. Display no answer and store the question in another file to check by the system administrator and to build a corpus electronically by different users.

1. First: the system compares between the query and the previous query in the database, if it found, then the system read the answer directly and display it.
2. Second: if there is no matching question in the database, then the system checks the answers, if there is any matching then the system display it.
3. The system saves in the file of the questions that have answers and tagged it with its answer.
4. If the question doesn't have an answer or then the system saves it in a file of questions that doesn't have an answer to check by administrators and scholars; is it right or wrong.

At last, two main files produced from users' questions by the system: the first one contains the questions that have answers, the second one contains the questions that don't have answers.

In the following parts illustrates the system work and some contents of the files by figures.

1. The system accepts the query first in a dialog box as in figure26:

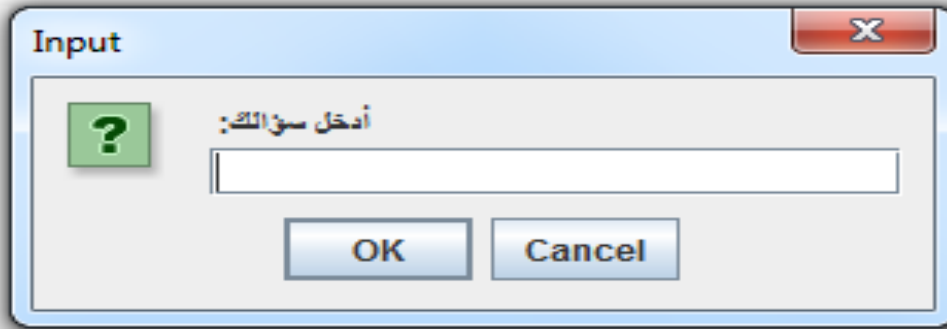


Figure 5—27: The system asked the user to enter its question.

1. The user insert the question as in figure [5-28]:

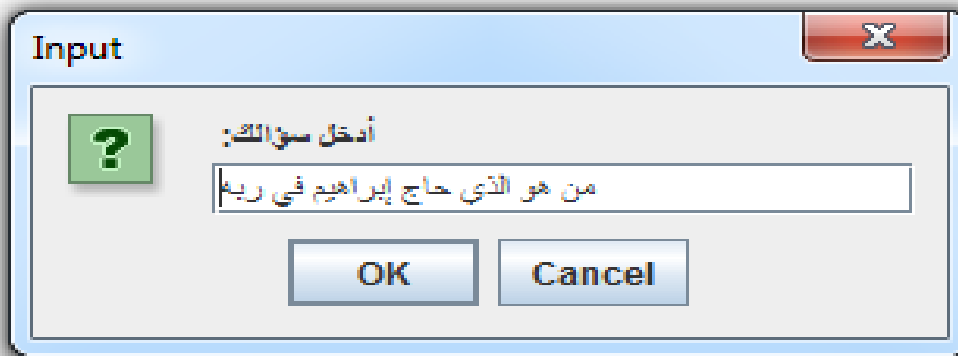


Figure 5—28:Accepting user query from the system.

2. The system displays the results as in figure [5-29]:

وَوَصَّىٰ بِهَا إِبْرَاهِيمَ بَنِيهِ وَيَعْقُوبَ يَا بَنِيَّ إِنَّ اللَّهَ اصْطَفَىٰ لَكُمُ الدِّينَ فَلَا تُمُونُوا إِلَّا وَآلَكُمْ مُسْلِمُونَ
 وَقَالُوا كُونُوا هُودًا أَوْ نَصَارَىٰ تَهْتَدُوا قُلْ بَلْ مِلَّةَ إِبْرَاهِيمَ حَنِيفًا وَمَا كَانَ مِنَ الْمُشْرِكِينَ
 قُلُوبُونَ إِنَّ إِبْرَاهِيمَ وَإِسْمَاعِيلَ وَيَعْقُوبَ وَالْأَسْبَاطَ كَانُوا هُودًا أَوْ نَصَارَىٰ قُلْ أَنتُمْ أَعْلَمُ أَمْ اللَّهُ وَمَنْ أَظْلَمُ مِمَّنْ كَتَمَ شَهَادَةَ عِندَهُ مِنَ اللَّهِ وَمَا اللَّهُ بِغَافِلٍ عَمَّا تَعْمَلُونَ
 ثُمَّ إِلَىٰ الَّذِي حَاجَّ إِبْرَاهِيمَ فِي رَبِّهِ أَنْ آتَاهُ اللَّهُ الْمُلْكَ إِذْ قَالَ إِبْرَاهِيمَ رَبِّيَ الَّذِي يُحْيِي وَيُمِيتُ قَالَ أَنَا أُحْيِي وَأُمِيتُ قَالَ إِبْرَاهِيمُ فَإِنَّ اللَّهَ يَأْتِي بِالشَّمْسِ مِنَ الْمَشْرِقِ فَأْتِ
 بِرَاهِيمَ رَبِّ أَرِنِي كَيْفَ تُحْيِي الْمَوْتَىٰ قَالَ أُولِمُ تُوْمِنُ قَالَ بَلَىٰ وَلَئِن لِّيَطْمِئِنَّ قُلُوبِي قَالَ فخذُ أَرْبَعَةً مِنَ الطَّيْرِ فَصُرْهُنَّ إِنَّكَ تَمُ اجْعَلْ عَلَىٰ كُلِّ جَبَلٍ مِنْهُنَّ جُزْءًا ثُمَّ ادْعُ
 فَتَلْهُنَّ أَتَمَّ مِنْ رَبِّهِ كَلِمَاتٍ فَتَابَ عَلَيْهِ إِنَّهُ هُوَ التَّوَّابُ الرَّحِيمُ
 قَالَ إِبْرَاهِيمَ رَبِّ اجْعَلْ هَذَا بَلَدًا آمِنًا وَارْزُقْ أَهْلَهُ مِنَ الثَّمَرَاتِ مَنْ آمَنَ مِنْهُم بِاللَّهِ وَالْيَوْمِ الْآخِرِ قَالَ وَمَنْ كَفَرَ فَأُمَتِّعُهُ قَلِيلًا ثُمَّ أَضْطَرُّهُ إِلَىٰ عَذَابِ النَّارِ وَبِئْسَ الْمَصِيرُ
 بِرَاهِيمَ رَبِّ أَرِنِي كَيْفَ تُحْيِي الْمَوْتَىٰ قَالَ أُولِمُ تُوْمِنُ قَالَ بَلَىٰ وَلَئِن لِّيَطْمِئِنَّ قُلُوبِي قَالَ فخذُ أَرْبَعَةً مِنَ الطَّيْرِ فَصُرْهُنَّ إِنَّكَ تَمُ اجْعَلْ عَلَىٰ كُلِّ جَبَلٍ مِنْهُنَّ جُزْءًا ثُمَّ ادْعُ

Figure 5—29:The result of the question in figure two

5.4.6.1 Challenges

Some challenges facing us, but the main problem is caused by the orthography of holy Quran, especially that symbols used as diacritics instead of the character such (pointed by arrow as in the following verse): (واو المد, ياء المد, ألف المد): that written in Holy Quran as in figure28, because the system doesn't display this symbol in a text file as in the following verse:



and most Muslims use Modern Standard Arabic (MSA). In the following we can If the user insert the following query: من هو الرحمن , then system search for the word الرحمن in the files to find the matching word, but the system doesn't find it because the word have another orthography shape, also this word appeared as the user writes it instead of the differences between the two words. The figures below describe this.

- The first figure below describes the answers in the text file and the word الرحمن underlined is different as it's written in a text file.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ
الرَّحْمَنِ الرَّحِيمِ
مَالِكِ يَوْمِ الدِّينِ
إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ
اهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ
صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ

Figure 5—30: The word الرحمن as its appeared in the file

- The answer of the above question presented as follows:

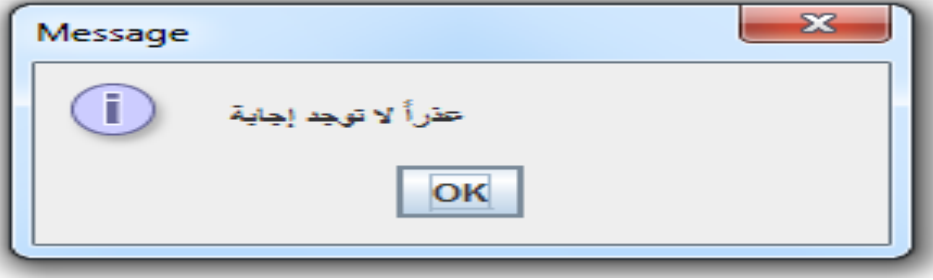


Figure 5—31: The answer of the above questions

If we change the word الرحمن in the text file the results appear as follow:

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
 الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ
 الرحمن الرحيم
 مَالِكِ يَوْمِ الدِّينِ
 إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ
 اهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ
 صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ

Figure 5—32: The changing of the word الرحمن in the text file.

- The system replies as in figure instead of three answers that displayed by other systems as in figure [5-33].

من هو الرحمن
 الرحمن الرحيم

Figure 5—33: The result of the question after changing the word

- Also the following figure displays the answer of the same question answered by another system which display three answers from the same answers.

Your Question is :

من هو الرحمن

اسم السورة البقرة
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
[In, the, name, of, Allah,, Most, Gracious,, Most, Merciful.]
رقم الآية 1

اسم السورة الفاتحة
الرَّحْمَنِ الرَّحِيمِ
[Most, Gracious,, Most, Merciful]
رقم الآية 3

اسم السورة البقرة
وَالْهَيْكُلُ إِلَهُ وَاحِدٌ لَا إِلَهَ إِلَّا هُوَ الرَّحْمَنُ الرَّحِيمُ
[And, your, Allah, is, One, Allah., There, is, no, god, but, He,, Most, Gracious,, Most, Merciful]
رقم الآية 163

عدد الإجابات = 3

Figure 5—34: The different answers displayed by another system.

5.4.6.2 The automatic corpus and its importance

The final result of this system is building two, corpora for the questions that answered, and not answered by the system. The following figure displays some examples of the questions.

من هو غبراهيم
من هو إبراهيم
من هو ابراهيم
من هو الرحمن الرحيم
من هو الرحمن
ما هو الحج
ما هي العمرة
ما هو الربا
من هم أشد الناس حباً لله
من هو الذي حاج إبراهيم في ربه

Figure 5—35: An example of automatic corpus for Quran questions

5.4.6.3 The automatic corpus importance

This corpus can be useful for many linguistic purposes, specially for Arabic language, which it can be used for spelling correction, frequent words, common mistakes in Arabic language, and many other features.

- **The prototype features:**

Depending on the first prototype this prototype has many features distinguish from it:

1. The main feature is to build corpuses directly, without any significant effort.
2. It can be applied in many domains (applied in the holy Quran and communications company for customer care domains).
3. Less code, although it has many new functions (less than 175 instead of 242 lines).

5.4.7 The fifth prototype (Enhanced prototype1 sing exaggeration formula)

5.4.7.1 Exaggeration formulas

In Arabic language Exaggeration formulas used to confirm the meaning and its repetition[113], and formula means the pattern of this word used to refer to exact meaning refer to it and no other formulas that meaning. Exaggeration formulas have five pattern in the Arabic language which are: **مَفْعَال**, **فَعِيل**, **فَاعِل**, **فَاعِل**, **فَاعِل** Fa'al, Fa'ol, Fa'el, Mifa'l, Fael. These pattern are very common in Arabic language, especially in the holy Quran, such as: Most Merciful **الرحيم** ، **الكريم** the generous, Knowing **العليم**, **الرؤوف** Most Gracious , **الشكور** The Grateful. For this we built our prototype to enhance the question answering system to solve this problem the following figure illustrates this.

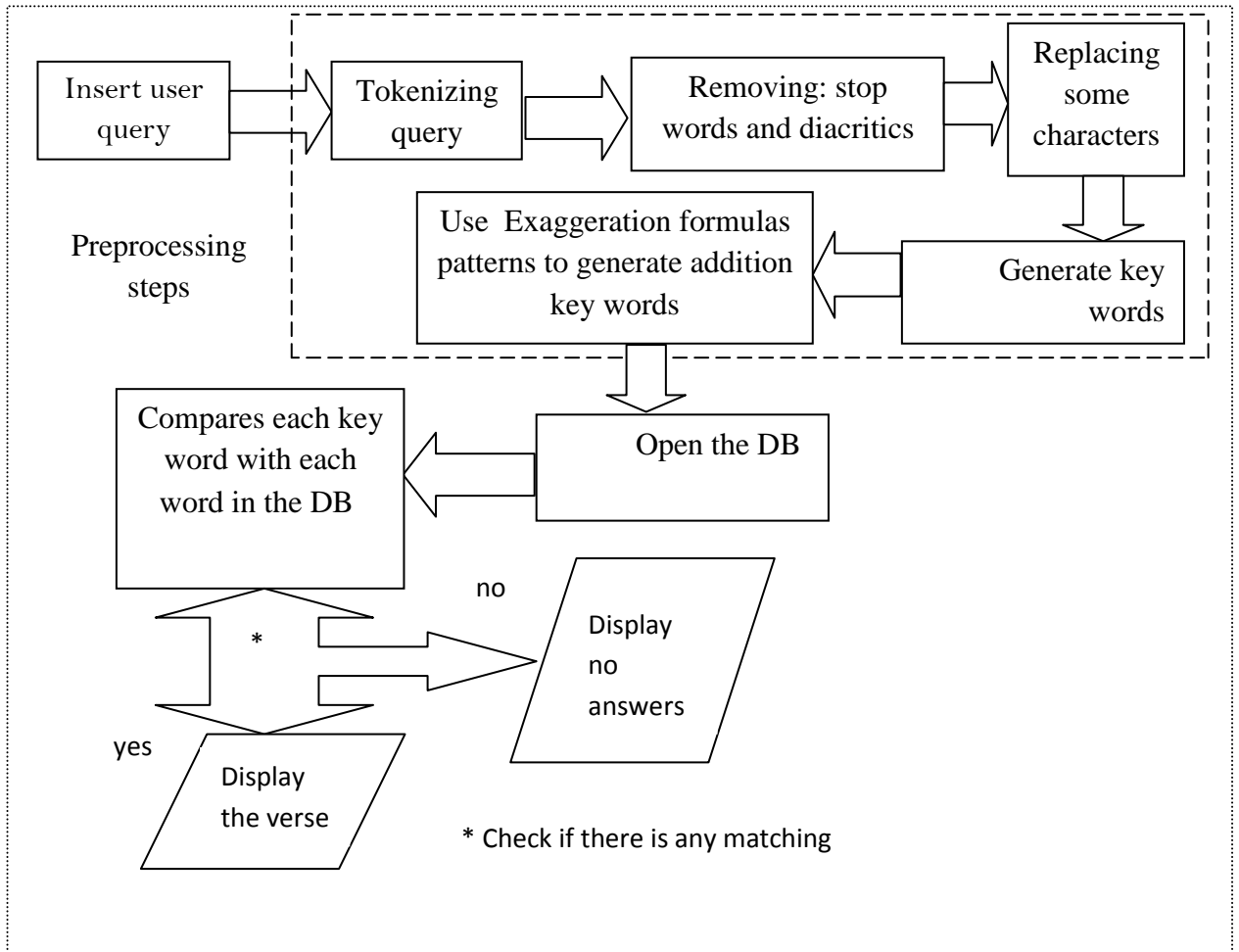


Figure 5-36 : prototypes 5 using Exaggeration formulas

Some details mentioned in the following parts to clarify the above processes and system works:

The pattern **فَعَّالٌ، فَعُولٌ، فَعِيلٌ** : first the system check the number of the characters of the word if 6, then check the first two numbers if the first is **ا** and the second is **ل** then remove them, and check the third one in the new word if it is **ا**, or **ي**, or **و** then change to the other two characters and save it. This done without any consideration to the diacritics.

5.4.7.2 Method used :

In general the system work as depicted in figure[5-36]. And following steps illustrated this.

1. Read the query from the user: the system read the query from user to process it.
2. Remove symbols and diacritics: the system removes some symbols and diacritics that affects in search such as ُ, َ, ِ .
3. The system tokenizes the questions (string of words) to splited word each one in string such as : من هو الرحمن الرحيم : the result of this step is four separated words: who من , هو هو , Most Gracious الرحمن , الرحيم Most Merciful .
4. Remove stop word: in this step the system remove the common words in the language such as: هو, من . The result of this step is the key word الرحمن , الرحيم .
5. The system use the key words to expand the query words by adding many pattern that match one of the formula exaggeration as mentioned below.
6. Classifying question: when the final key word prepared then the system check if the key words match one of the question types of Arabic language presented by [114] the the system choose the suitable corpus for this question, else the search the answers.
7. The system displays the verses matches the key words, else displays no no answer.

5.4.7.3 Experiments

Our experiments depend on a corpus and question answering system, which presented in [5], also this enhancement depends on the prototype that proposed.

Some details mentioned in the following parts:

1. The weight فَعِيل ، فَعُول ، فَعَّال : first the system check the number of the characters of the word if 6, then check the the first two numbers if the first is ا and the second is ل then remove them, and check the the third one in the new word if it is ا , or ي , or و then change to the other two characters and save it. This done without any consideration to the diacritics. The following code explains this:
2. The weight مَفْعَال : first the system check the number of the characters of the word if 5, then check the the first one if the first is م and the fourth one is ل then, and check the the third one in the new word if it is ا , or ي , or و then change to the other two characters and save it. This done without any consideration to the diacritics. The following code explains this:

5.4.7.4 Some examples and comparisons:

Here are some examples and results done by the system in the:

1. First the system accepts the question from the users, if we insert the query : من هو الرحيم the system removes the word من and the word هو, then the word الرحيم formulated to the following words: رحيم، رحوم، رحام، الرحيم, then the system displays the answers depending on the matching words.

سؤالك هو :
من هو الرحيم

سؤالك هو بعد تعديل الكلمات الأساسية في البحث :
من هو الرحيم رحام رحوم رحيم

اسم السورة البقرة
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
[In, the, name, of, Allah,, Most, Gracious,, Most, Merciful.]
رقم الآية 1

اسم السورة البقرة
رَقْمُ الْآيَةِ 218
إِنَّ الَّذِينَ آمَنُوا وَالَّذِينَ هَاجَرُوا وَجَاهَدُوا فِي سَبِيلِ اللَّهِ أُولَئِكَ يَرْجُونَ رَحْمَةَ اللَّهِ وَاللَّهُ غَفُورٌ رَحِيمٌ
[Those, who, believed, and, those, who, suffered, exile, and, fought, (and, strove, and, struggled)]

اسم السورة البقرة
رَقْمُ الْآيَةِ 226
لِلَّذِينَ يُؤْلُونَ مِنْ نِسَائِهِمْ تَرَبُّصًا أَرْبَعَةَ أَشْهُرٍ فَإِنْ فَاءُوا فَإِنَّ اللَّهَ غَفُورٌ رَحِيمٌ
[For, those, who, take, an, oath, for, abstention, from, their, wives,, a, waiting, for, four, months,

عدد الإجابات = 11

Total time =11

BUILD SUCCESSFUL (total time: 14 seconds)

Figure 5—36: Answers displayed by prototype 5

From figure: [5-36] we notice that the number of results here becomes 11 answers instead of 5 answers as in table 10, which explains that there is a good improving in results.

Also a comparison between prototype one and five depending on table 10 and the following formula:

Percentage = number of right or wrong answer x 100%

Total number of answers

Table 5—14: comparison between prototype one and five

Question	Prototype 1			Prototype 5		
	No of answers	Related & %	Wrong & %	No of answers	Related & %	Wrong & %
من هو الرحمن الرحيم Who is the Most Gracious and . Most Merciful	5	45.5	54.5	11	100	0
من هم بني اسرائيل Who are the children of Israel	7	85.7	14.3	6	100	0
ما هو الصيام What is fasting	2	66.6	33.3	3	100	0
من هو الحي القيوم Who is living and subsisting	1	100	0	1	100	0

5.4.8 Prototype 6 singular, dual, and plural

These experiments depend on theoretical linguistic features of the Arabic language specifically concerning nouns. In the Arabic language the noun class is divided into sub-categories as in the figure[5-38] below:

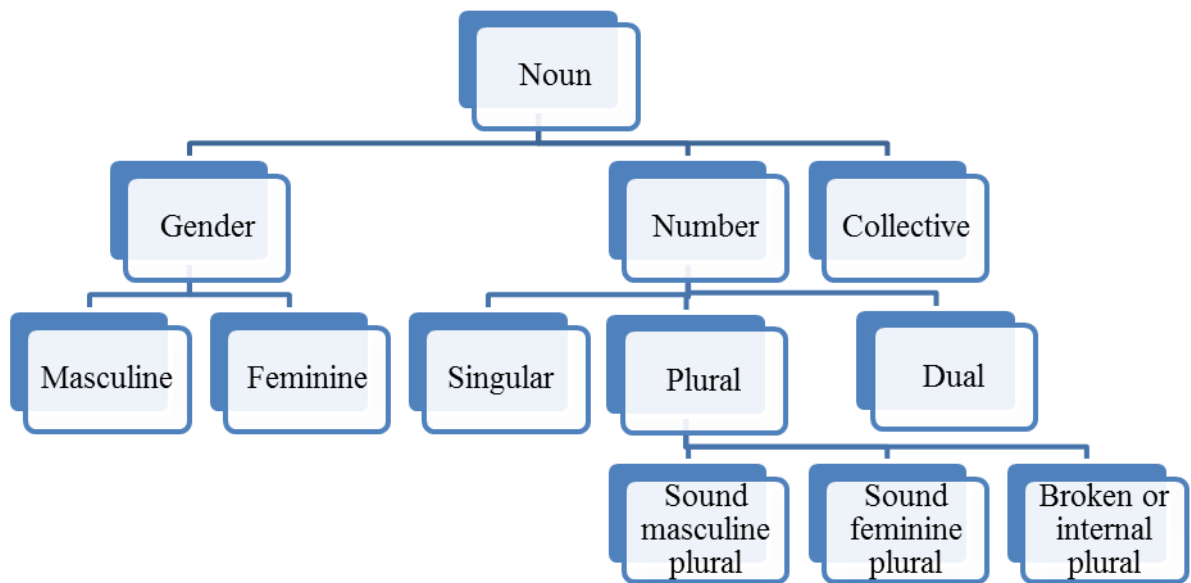


Figure 5—37: Noun category in Arabic language

Depending on this noun category-set, we built our prototype, which deals with the word: if the word is in the singular, then replace it by a plural, or dual and vice versa. The following part explains the nounnumber system in Arabic:

- Singular is used to indicate one person or thing, the Arabic word here must be abstracted from any mark for dual or plural, such as: ولد (boy), كرة (foot ball).
- Dual is used to indicate two persons or things, such as: ولدان (boy), كرتين (foot ball). The indicator here is: ان, ين depending on its position in the sentence.
- Plural: in the plural parts there is some complexity, because plural is divided into sub-categories: regular plurals or sound plurals, irregular plurals or broken plurals, and the collective. Sound plural are also divided into two sub-categories: masculine and feminine. The broken plural changes the singular weight in irregular changes without fixed rules, where fixed rules are used in sound masculine plural (adding: ون, ين at the end of the singular) and sound feminine plural (adding : ات at the end of the singular). The collective plural, which is called a noun plural, is used as a

singular to explain the plural by deleting the feminine mark (ة) or ي (called ya alnasab (ياء النسب)). Table [5-15] gives some examples to clarify this.

Table 5-15: Examples of noun number features in the Arabic language

ملحوظات	نوعه type	علامته mark	الجمع	علامته mark	المثنى dual	نوعه	المفرد singular
	سالم	ون / ين	معلمون / معلمين	ان / ين	معلمان / معلمين	مذكر	معلم
	سالم	ات	معلمات	ان	معلمتان	مؤنث	معلمة
	سالم	ات	برتقالات	ان	برتقالتان	مؤنث	برتقالة
	اسم جنس جمعي	حذف الة	برتقال	ان	برتقالتان	مؤنث	برتقالة
	اسم جنس جمعي	حذف الـ ي	جند	ان	جنديان	مذكر	جندي
إضافة حروف للمفرد	تكسير	لاعلامه	أولاد	ان	ولدان	مذكر	ولد
نقصان حرف من المفرد	تكسير	حذف الألف	كتب	ان	كتابان	مذكر	كتاب
تغيير في أصل الحرف	تكسير	تغيير حرف	دور	ان	داران	مذكر	دار
إضافة وتغيير	تكسير	إضافة وتغيير	قواميس	ان	قاموسان	مذكر	قاموس

5.4.8.1 The theoretical model

Depending on the above outline description of the Arabic language noun system, characteristics of Arabic noun number used to extend the Question-Answering system. The extension builds patterns from the words entered depending on three

weights (4,6,8); these are general and used in the Holy Quran frequently. Then check its weight: if the weight is found then apply the rule of this weigh as in the following algorithms:

```

1. Check the weight of the word :
   if (6 or 8) then apply the following :
   originalword = acceptedword
   if the first 2 characters of the acceptedword are ال then remove it

for I = 1 to 8

begin

finalword= the original word

newword = the original word+ ان Or newword = ال +the original word+ان;

finalword= Finalword+ newword

newword = the original word+ ات Or newword = ال +the original word+ات;

finalword= Finalword+ newword

newword = the original word+ ون Or newword = ال +the original word+ون;

finalword= Finalword+ newword

newword = the original word+ ين Or newword = ال +the original word+ين;

end;

```

5.4.8.2 The practical implementation of the system

As in figure[5-36] used in prototype 5 ,Exaggeration formulas pattern substituted with singular, dual and plural with or without ال, also a corpus of Quran questions and answers was used to answer the questions entered by the users to test the system. Only one question here is shown as an example. Finally we carried out a comparison between original and extended systems.

The system works as follows:

1. The system first accepts the query “Who is the believer?” من هو المؤمن from the user as in figure [5-39]:

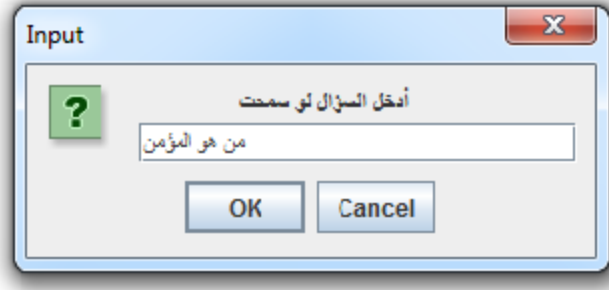


Figure 5—38: user query من هو المؤمن

The system displays the results as in figure[5-40].

سؤالك هو بعد تعديل الكلمات الأساسية في البحث :

المؤمن مؤمن مؤمنان مؤمنون مؤمنات مؤمنين المؤمنان المؤمنون المؤمنات مؤمنان مؤمنات مؤمنين مؤمنات مؤمنين

اسم السورة البقرة رقم الآية 91
 وإذا قيل لهم آمنوا بما أنزل علينا ويكفرون بما وراءه وهو الحق مصدقا لما معهم قل ولم نقولن أنبياء الله من قبل إن كنتم مؤمنين
 [When, it, is, said, to, them,, "Believe, in, what, Allah, Hath, sent, down,, "they, say,, "We, believe, in, what, was, sent, down, to, us:", yet, t]

اسم السورة البقرة رقم الآية 221
 ولا تتكفوا المشركيات حتى يؤمنن ولأمة مؤمنة خير من مشركية ولو أحببتكم ولا تتكفوا المشركين حتى يؤمنوا ولعبدا مؤمنا خيرا من مشرك ولو أحببتكم أولئك يدعون إلى ال
 [Do, not, marry, unbelieving, women, (idolaters),, until, they, believe:, A, slave, woman, who, believes, is, better, than, an, unbelieving, wor]

اسم السورة البقرة رقم الآية 223
 يسألوكم عن حشركم فقولوا نعم إنكم فائتوا حشرهم ألى سيئتهم وقدموا لأنفسكم واتقوا الله واعلموا أنكم ملائكة وبشر المؤمنين
 [Your, wives, are, as, a, tilth, unto, you,, so, approach, your, tilth, when, or, how, ye, will:, but, do, some, good, act, for, your, souls, beforeh]

اسم السورة البقرة رقم الآية 248
 وقال لهم نبيهم إن آية ملكه أن يأتيكم التابوت فيه سكينه من ربكم وبقية مما ترك آل موسى وآل هارون تحمله الملائكة إن في ذلك لآية لكم إن كنتم مؤمنين
 [And, (further), their, Prophet, said, to, them:, "A, Sign, of, his, authority, is, that, there, shall, come, to, you, the, Ark, of, the, covenant,, with]

اسم السورة البقرة رقم الآية 278
 يا أيها الذين آمنوا اتقوا الله وذرُوا ما بقِي من الربَا إن كنتم مؤمنين
 [O, ye, who, believe!, Fear, Allah,, and, give, up, what, remains, of, your, demand, for, usury,, if, ye, are, indeed, believers]

عدد الإجابات = 5

Figure 5—39: answer to the user query من هو المؤمن

5.4.8.3 Example question and answers

1. The system searches for the query used by the user
2. The query expansion generates the expanded query: مؤمن، المؤمن، المؤمنان، المؤمنات، المؤمنون، المؤمنات، “Who is the believer, the believer, the believers, the believers, the believers” which is one word that uses many patterns.

3. First the system displays the new pattern for only one word the المؤمن generate the following patterns (new 14 words) for search as in table [5-16]:

Table 5—15: the result word (patterns) of the query and some remarks

ملاحظات	نوعها من حيث التذكير و التأنيث		نوعها من حيث العدد			الكلمة The word
	Gender		Type in terms of number			
معرف بال / نكرة	مؤنث	مذكر	جمع	مثنى	مفرد	
Definite Noun with AL / indefinite	feminine	masculine	plural	dual	singular	
نكرةindefinite		√			√	مؤمن believer
نكرةindefinite	√				√	مؤمنات believers
نكرةindefinite	√			√		مؤمنان Two believers
نكرةindefinite		√		√		مؤمنين Two believers
نكرةindefinite		√	√			مؤمنون believers
معرف بال / Definite		√			√	المؤمن

						believer
/ معرف بال Definite	√		√			المؤمنات believers
/ معرف بال Definite		√		√		المؤمنين Two believers
/ معرف بال Definite		√	√			المؤمنون Believers
/ معرف بال Definite		√		√		المؤمنان Two believers
/ معرف بال Definite	√			√		المؤمنتان Two believers
نكرة indefinite	√			√		مؤمنتين Two believers
نكرة indefinite	√			√		مؤمنتان Two believers
/ معرف بال Definite	√			√		المؤمنتين Two

						believers
--	--	--	--	--	--	-----------

4. The second process is removing stop words, some symbols, and diacritics which affect the search results.
5. The system uses the generated patterns to make a match between the patterns and the corpus. If any match is found then store the results; if there is not any new answers then display the results.

5.4.8.4 Comparison between baseline and extended systems

The following section compares a baseline Quran QA system, and the new extension.

First we apply the same question; the answer is :

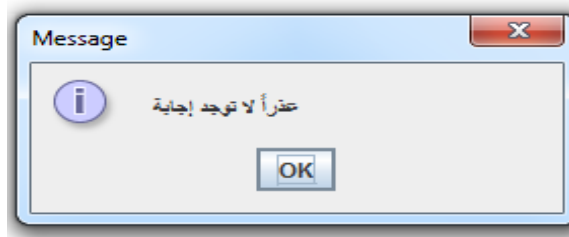


Figure 5—40: the answer of the question من هو المؤمن

When the question is : من هم المؤمنین, the answer is من هم المؤمنین, only one answer appeared as appeared in figure [39]:

Your Question is :

المؤمنين

اسم السورة رقم الآية رقم السورة
المؤمنين

2 البقرة 223
يَسْأَلُكُمْ خِزْيَانُ لَكُمْ فَاَتُوا خِزْيَانَكُمْ اَنَّى شِئْتُمْ وَقَدِّمُوا لِانْفُسِكُمْ وَاتَّقُوا اللَّهَ وَاعْلَمُوا اَنَّكُمْ مُلَاقُوهُ وَبَشِّرِ الْمُؤْمِنِينَ:الآية
عدد الإجابات = 1
BUILD SUCCESSFUL (total time: 9 seconds)

Figure 5—41: the answer of the question من هم المؤمنین

We notice there is only one answer because it depends on the key word matching techniques, instead of 5 answers as appeared in the extended system.

5.4.8.5 Experiments and results

In this part we discuss our experiments and results, which were judged by Islamic scholars from Gabrah college. We used 30 questions from our Quran question answering corpus applied in the systems. The following tables and figure explain the results.

Table 5—16: General table results

Question Number	QA (1)		QA (2)	
	Right answer	Matching	Right answer	Matching
2	1	1	1	1
12	0	3	7	7
13	0	0	2	2
16	1	1	2	2
22	1	1	2	4
31	1	8	8	8
32	1	29	13	30
76	1	4	6	8
91	0	0	1	6
110	1	7	4	7
111	1	31	1	32
119	1	4	1	7
121	1	8	1	10

132	1	6	1	6
137	1	2	1	2
165	1	1	1	5
168	0	7	0	10
179	1	16	0	6
187	1	1	2	2
188	0	1	0	2
199	0	4	0	9
201	0	0	1	2
214	0	0	1	73
226	1	5	1	36
238	1	9	3	8
239	1	6	1	6
241	0	4	0	1
251	1	13	1	16
275	1	13	1	4
281	1	8	3	23
Summation	21	193	66	355

From table [5-18] depending on the final summation the following tables generated depending on the following formula:

$$\text{Percentage} = \frac{\text{number of right or wrong answer}}{\text{Total number of answers(of right or wrong answer)}} \times 100\%$$

Total number of answers(of right or wrong answer)

Total of right answers= 71

QA	QA1	QA2	Differences
No of questions	30	30	0
No of right answers	21	66	45
No of wrong answers	50	5	45
% right	29.5	92.9	45.2
% wrong	70.5	7.04	35.2
% wrong (matching)	25.9	81.4	4.1
% right (matching)	10.8	18.3	7.7

Figure 5—42:Comparative table for results

Also from table 4 we can generate the following chart as in figure [5-42]:

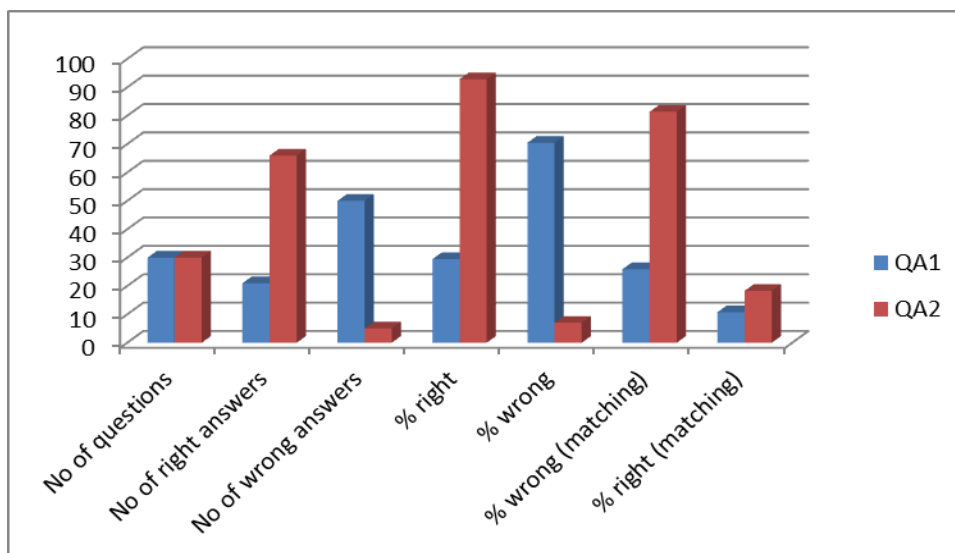


Figure 5—43: comparison between prototype 1 and 6

From all the above experiments we notice that:

1. The increase of right answer in extended QA 2.
2. The decrease of wrong answers in extended QA 2.

5.4.9 Experiments Summary

Six prototypes were built, each one has its characteristics. The following table illustrates this in details:

Table 5—17: Prototypes summarizations

Characteristics Prototype No	NLP processes	Corpus	Pattern used	Storage means	Notices
1	Stop word, diacritics, tokenizing, key word matching	Dynamic corpus used	Not used	Database	Less stop words
2	Stop word, diacritics, tokenizing, key word	Dynamic corpus used	One pattern used	Text files	Lucene search engine

	matching		فياعيل		used
3	Stop word, diacritics, tokenizing, key word matching	Dynamic corpus used	، فياعيل ، تفاعيل ، فياعيل	Text files	Lucene search engine used
4	Stop word, diacritics, tokenizing, key word matching	Used and using the users query to build dynamic corpus	Not used	Text files	
5	stop word, diacritics, tokenizing, key word matching	Dynamic corpus used	Eight patterns used ، فياعيل، تفاعيل، فياعيل and exaggeration formulas	Database	More addition stop word and more aptterns
6	stop word, diacritics, tokenizing, key word matching	Dynamic corpus used	Many patterns used using singular, dual, plurals in Arabic numbers system	Database	More addition stop word and more aptterns

Table 5—18: Prototypes summarizations

Depending on [115] many criteria 's to evaluate computational Quran search. The following table describes the 13 criteria's (that used to compare between many application and tools) to be applied in question answering system prototypes. The following following illustrate this.

1. Search techniques : it used semantic and Keywords Search.
2. Query analyzer: Stem, and Lemma of query words.
3. Quranic Ontologies: not used.

4. No. of Ontology: Zero.
5. Quran Datasets: part of Holy Quran (Alfatihah and AlBagarah Chapters), and some questions.
6. Number of dataset types: Three types of data set : questions, Holy Quran verses (Alfatihah and AlBagarah Chapters), and translations of Holy Quran verses (Alfatihah and AlBagarah Chapters) by Abdullah Yusuf Ali [36].
7. Query types: one word, two words, sentences, and questions.
8. Results types: Verse.
9. Availability: not available.
10. Result Ranking: non ranked.
11. Systems Users: general users.
12. Search Domain Coverage: part of Holy Quran verses (Alfatihah and AlBagarah Chapters).
13. Language of query: Arabic and English language.

5.5 Chapter summarization

In this chapters, many experiments were done, a corpus built from two chapters of the Holy Quran verses and real question collected from users and internet, six prototypes were building each one have some characteristics different from others, and discussing results, lastly comparative table and evaluation criterias were applied.

CHAPTER VI

6 Chapter six: Conclusion and recommendation

This chapter concludes the thesis and summarize the contribution of this thesis and the direction of future works.

6.1 Conclusions

Applying question answering system in the holy Quran field is very important, because most of the Muslims need to search about some verses to find judgment or knowledge. In this study Question answering system for holy Quran applied for only two chapters Bagrah and Fatihah (البقرة والفاطحة). A corpus of verses and chapters was built (only AlBagrah and Alfatihah chapters), and their translations. Also corpus of questions collected and revised by scholars from the Gabrah Academic College. Six prototypes depending on many characteristics (key word match and removing stop words, indexing and using Broken plural, exaggeration formula, singular, dual, and sound plural, and extracting and ranking answers). The the corpus applied on these prototypes, varies in results, but in general each prototype enhances the results: recall and precision. Designing these prototypes and their methods help computation linguistics researcher to adapt their systems to enhance query expansions and the question answering system results in precision and recall. The thesis contains six chapters, these chapters structured as follows:

Chapter one defines the problem statement, objectives and motivation, and research methodology.

Chapter two gives a general introduction and history to question answering system , natural language processing, and corpus.

Chapter three reviews the related work in many fields: Arabic language, Holy Quran, Corpus, question answering systems.

Chapter four important introductions to Arabic language and its characteristics, Holy Quran characteristics and its role in safe Arabic language.

Chapter five contains methodology, methods, experiments 6 experiments were done (first one uses base NLP question answering system removing stopwords,

diacritics, and some symbols, the second uses indexing and تفاعيل pattern, the third uses indexing and فياعيل، فعاعيل، تفاعيل pattern, the fourth uses dynamic corpus build by real user question, the fifth uses Exaggeration formulas pattern, the sixth uses singular, dual and plural pattern), results and discussions.

6.2 Recommendations for further research

1. Applying all the patterns in one prototype to add more enhances to the question answering system for Arabic language.
2. Applying these prototypes in other domains, such as Hadith, Tafseer and Arabic Bible.
3. Use an ontology for semantic analysis in matching questions.

7 References

- [1] P. Kumar, S. Kashyap, A. Mittal, and S. Gupta, 'A Fully Automatic Question-Answering System for Intelligent Search in E-Learn...', *International Journal on E-Learning*, vol. 4, pp. 149–166, 2005.
- [2] S. J. Athenikos and H. Han, 'Biomedical question answering: a survey.', *Computer methods and programs in biomedicine*, vol. 99, no. 1, pp. 1–24, Jul. 2010.
- [3] V. Gupta, 'A Proposed Online Approach of English and Punjabi Question Answering', *International Journal of Engineering Trends and Technology (IJETT)*, vol. 6, no. 5, pp. 292–295, 2013.
- [4] D. R. Radev, H. Qi, Z. Zheng, S. Blair-Goldensohn, Z. Zhang, W. Fan, and J. Prager, 'Mining the web for answers to natural language questions', *Proceedings of the tenth international conference on Information and knowledge management - CIKM'01*, p. 143, 2001.
- [5] M. a Bauer and D. Berleant, 'Usability survey of biomedical question answering systems.', *Human genomics*, vol. 6, p. 17, Jan. 2012.
- [6] I. B. I. Ebtihal Mustafa AlAmin, Abbakar Musa Ali, Maha Ali Mahmoud, 'IDENTIFYING ARABIC BROKEN PLURAL IN INFORMATION RETRIEVAL SYSTEM', Sudan University of Science and Technology, 2015.
- [7] M. T. M. Mahmoud, 'Designing Broken Plurals Processing Method for Enhancing the Performance of Arabic Information Retrieval Systems', Sudan University of Science and Technology, 2015.
- [8] and E. L. B. Steven Bird, Ewan Klein, *Natural Language Processing with Python*, First. O'Reilly Media, 2009.
- [9] M. Hammo, B., Sleit, A., El-Haj, 'Effectiveness of Query Expansion in Searching the Holy Quran', *colloque internationale Traitement automatique de la langue Arabe.*, CITALA, vol. 7, pp. 18–19, 2007.
- [10] R. Te, 'Research on Question Classification Method of Tibetan Online Automatic Question-Answering System', *2011 4th International Conference on Intelligent Networks and Intelligent Systems*, pp. 211–213, Nov. 2011.
- [11] L. Hirschman and R. Gaizauskas, 'Natural language question answering: the view from here', *Natural Language Engineering*, vol. 7, no. 4, pp. 275–300, Feb. 2002.
- [12] T. Kalaiyarasi, R. Parthasarathi, and T. V Geetha, 'P O O N G K U Z H A L I - A n I n t e l l i g e n t T a m i l C h a t t e r b o t', 1950.
- [13] W. a. Woods, 'Progress in natural language understanding: an application to lunar geology', *Proceedings of the National Computer Conference and Exposition on AFIPS '73*, pp. 441–450, 1973.
- [14] B. F. Green, A. K. Wolf, C. Chomsky, and K. Laughery, 'BASEBALL: an automatic question answerer', *Proceedings of western joint IRE-AIEE-ACM '61 computer conference*, pp. 219–224, 1961.
- [15] R. F. Simmons, 'Answering English questions by computer: a survey',

- Communications of the ACM*, vol. 8, no. 1, pp. 53–70, 1965.
- [16] R. F. Simmons, ‘Natural language question-answering systems: 1969’, *Communications of the ACM*, vol. 13, no. 1, pp. 15–30, Jan. 1970.
- [17] M. Pacak and A. W. Pratt, ‘The function of semantics in automated language processing’, *Proceedings of the 1971 international ACM SIGIR conference on Information storage and retrieval - SIGIR '71*, 1971.
- [18] D. L. Waltz, ‘An English language question answering system for a large relational database’, *Communications of the ACM*, vol. 21, no. 7, pp. 526–539, 1978.
- [19] A. C. Graesser, S. E. Gordon, and L. E. Brainerd, ‘QUEST: A model of question answering’, *Computers & Mathematics with Applications*, vol. 23, no. 6–9, pp. 733–745, 1992.
- [20] J. K. Xerox, J. K. Xerox, G. P. H. A. Road, G. P. H. A. Road, R. P. Alto, R. P. Alto, C. Ca, and C. Ca, ‘MURAX: A Robust Linguistic Approach For Question Answering Using An On-Line Encyclopedia’, *Components*, pp. 181–190, 1993.
- [21] Ó. Ferrández, R. Izquierdo, S. Ferrández, and J. L. Vicedo, ‘Addressing ontology-based question answering with collections of user queries’, *Information Processing and Management*, vol. 45, no. 2, pp. 175–188, 2009.
- [22] S. T. Pham and D. T. Nguyen, ‘RESOLVING THE SEMANTICS OF VIETNAMESE QUESTIONS IN VNEWSQA/ICT SYSTEM’, vol. 5, no. 2, pp. 11–20, 2014.
- [23] N. Haque, ‘A prototype framework for a Bangla question answering system using translation based on transliteration and table look-up as an interface for the medical domain April 2010 University of Malta’, 2010.
- [24] A. Mohamed, N. Allam, and M. H. Haggag, ‘The Question Answering Systems : A Survey .’, vol. 2, no. 3, 2012.
- [25] S. Kalaivani and K. Duraiswamy, ‘Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment’, *Journal of Computer Science*, vol. 8, no. 9, pp. 1407–1413, 2012.
- [26] H. Abdelnasser, ‘Al-Bayan : An Arabic Question Answering System for the Holy Quran’, *Proceedings of the 9th International Workshop on Semantic Evaluation*, pp. 57–64, 2014.
- [27] C. D. Paice, ‘An Evaluation Method for Stemming Algorithms’, *In Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 42–50, 1994.
- [28] G. Kanaan, A. Hammouri, R. Al-Shalabi, and M. Swalha, ‘A New Question Answering System for the Arabic Language’, *American Journal of Applied Sciences*, vol. 6, no. 4, pp. 797–805, Apr. 2009.
- [29] M. a Yunus, R. Zainuddin, and N. Abdullah, ‘Semantic query for Quran documents results’, *2010 IEEE Conference on Open Systems (ICOS 2010)*, no. Icos, pp. 1–5, Dec. 2010.

- [30] M. N. Al-Kabi, 'Towards improving Khoja rule-based Arabic stemmer', in *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies, AEECT 2013*, 2013, pp. 1–6.
- [31] A. Al-Omari and B. Abuata, 'Arabic light stemmer (ARS)', *Journal of Engineering Science and Technology*, vol. 9, no. 6, pp. 702–717, 2014.
- [32] N. Schlaefer, P. Gieselmann, T. Schaaf, and A. Waibel, 'A pattern learning approach to question answering within the Ephyra framework', *Text, Speech and Dialogue, Proceedings*, vol. 4188, pp. 687–694, 2006.
- [33] N. J. Belkin, 'Interaction with texts: Information retrieval as information-seeking behavior', *Information retrieval '93: von der Modellierung zur Anwendung Proceedings of the First Conference of the Gessellschaft für Informatik Fachgruppe Information Retrieval*, pp. 55–66, 1993.
- [34] K. Jbara, 'Knowledge Discovery in Al-Hadith Using Text Classification Algorithm', vol. 6, no. 11, pp. 485–494, 2010.
- [35] A. M. N. Allam and M. H. Haggag, 'The Question Answering Systems : A Survey', *International Journal of Research and Reviews in Information Sciences (IJRRIS)*, vol. 2, no. 3, pp. 211–221, 2012.
- [36] R. Srihari and W. Li, 'A question answering system supported by information extraction', *Proceedings of the sixth conference on Applied natural language processing -*, pp. 166–172, 2000.
- [37] R. J. Mooney and R. Bunescu, 'Mining knowledge from text using information extraction', *ACM SIGKDD Explorations Newsletter*, vol. 7, no. 1, pp. 3–10, 2005.
- [38] M. Wang, 'A Survey of Answer Extraction Techniques in Factoid Question Answering', *association of computational linguistics*, vol. 1, 2006.
- [39] A. Téllez-Valero, M. Montes-Y-Gómez, L. Villaseñor-Pineda, and A. P. Padilla, 'Learning to select the correct answer in multi-stream question answering', *Information Processing and Management*, vol. 47, no. 6, pp. 856–869, 2011.
- [40] E. Breck, J. D. Burger, L. Ferro, L. Hirschman, D. House, M. Light, and I. Mani, 'How to Evaluate your Question Answering System Every Day and Still Get Real Work Done', p. 6, 2000.
- [41] A. Allam and M. Haggag, 'The Question Answering Systems: A Survey', *International Journal of research and Reviews in Information Sciences, IJRRIS*, vol. 2, no. 3, 2012.
- [42] M. Paşca, 'Processing Definition Questions in an Open-Domain Question Answering System', in *Proceedings of the 2002 AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases*, 2002, pp. 73–78.
- [43] S. Quarteroni and S. Manandhar, 'Designing an interactive open-domain question answering system', *Natural Language Engineering 1*, vol. 1, no. 1, pp. 1–23, 2009.
- [44] G. Sureshkumar and G. Zayaraz, 'Domain Ontology Construction using

- Automatic Concept Relation Learning for Question Answering Systems’, *International Journal of Information and Data Management (IJIDM)*, vol. 1, no. February, pp. 9–15, 2013.
- [45] U. Andersen, A. Braasch, L. Henriksen, C. Huszka, A. Johannsen, B. Maegaard, O. Norgaard, S. Schulz, and J. Wedekind, ‘Creation and use of language resources in a question-answering e-health system’, in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, 2012, pp. 2536–2542.
- [46] P. Kumar, S. Kashyap, a. Mittal, and S. Gupta, ‘A Hindi Question Answering system for E-learning documents’, *2005 3rd International Conference on Intelligent Sensing and Information Processing*, pp. 80–85, 2005.
- [47] S. Mittal, S. Gupta, and A. Mittal, ‘BionQA Multidocument Question Answering System: Providing Access to E-learning for masses’, *Journal of Engineering Students, CAFFET INNOVA Technical Society, India*, pp. 28–37, 2008.
- [48] S. Cruchet, A. Gaudinat, C. Boyer, A. Morandi, and D. Marino, ‘Multilingual Question/Answering System applied to trusted health information’, in *Proceedings of the 22nd International Conference of the European Federation for Medical Informatics (MIE)*, 2009.
- [49] D. A. N. Moldovan and M. Pas, ‘Performance Issues and Error Analysis in an Open-Domain Question Answering System’, vol. 21, no. 2, pp. 133–154, 2003.
- [50] J. Chu-carroll, D. Ferrucci, J. Prager, and C. Welty, ‘Hybridization in Question Answering Systems’, *AAAI*, 2003.
- [51] D. Lenat, ‘CYC: A Large-Scale Investment in Knowledge Infrastructure’, *Communications of the ACM*, vol. 38, no. 11, pp. 33–38, 1995.
- [52] G. A. Antonio, A. Cisternino, F. Formica, M. Simi, and R. Tommasi, ‘PiQASso: Pisa Question Answering System’, in *Proceedings of the Text REtrieval Conference*, 2001, pp. 599–607.
- [53] S. Granger, ‘The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies’, *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, pp. 17–29, 2003.
- [54] B. A. Shawar and E. Atwell, ‘A chatbot system as a tool to animate a corpus’, pp. 5–24, 1999.
- [55] W. J. Plath, ‘REQUEST: A Natural Language Question-Answering System’, *IBM Journal of Research and Development*, vol. 20, no. 4, pp. 326–335, 1976.
- [56] J. Fukumoto, N. Aburai, and R. Yamanishi, ‘Interactive Document Expansion for Answer Extraction of Question Answering System’, *Procedia Computer Science*, vol. 22, pp. 991–1000, Jan. 2013.
- [57] C. Meenakshi, P. Thangaraj, and M. Ramasamy, ‘A Novel Scheme to Identify the Word Sense in Question Answering Systems’, vol. 2, no. 9, 2011.
- [58] G. Suresh kumar and G. Zayaraz, ‘Concept relation extraction using Naïve

- Bayes classifier for ontology-based question answering systems’, *Journal of King Saud University - Computer and Information Sciences*, May 2014.
- [59] L. Zhenqiu, ‘Design of Automatic Question Answering System Base on CBR’, *Procedia Engineering*, vol. 29, pp. 981–985, Jan. 2012.
- [60] B. Hammo, S. Abuleil, S. Lytinen, and M. Evens, ‘Experimenting with a question answering system for the Arabic language’, *Computers and the Humanities*, vol. 38, no. 4, pp. 397–415, Nov. 2004.
- [61] W. Brini, M. Ellouze, O. Tirgui, S. Mesfar, L. Hadrich Belguith, and P. Rosso, ‘Factoid and definitional Arabic Question Answering system’, *Arabian Journal for Science and Engineering 396 (2014): 4541-4564*, vol. 51, no. 4, pp. 769–784, 2009.
- [62] W. N Bdour and N. K. Gharaibeh, ‘Development of Yes/No Arabic Question Answering System’, *International Journal of Artificial Intelligence & Applications*, vol. 4, no. 1, pp. 51–63, Jan. 2013.
- [63] M. F. Noordin and R. Othman, ‘An Information Retrieval System for Quranic Texts: A Proposed System Design’, *2006 2nd International Conference on Information & Communication Technologies*, vol. 1, pp. 1704–1709, 2006.
- [64] M. Sharaf and E. S. Atwell, ‘QurAna: Corpus of the Quran annotated with Pronominal Anaphora.’, in *8th International Conference on Language Resources and Evaluation (LREC’12)*, 2012, pp. 130–137.
- [65] A. R. Yauri, R. A. Kadir, A. Azman, and M. A. A. Murad, ‘Ontology semantic approach to extraction of knowledge from holy quran’, *2013 5th International Conference on Computer Science and Information Technology*, pp. 1–5, Mar. 2013.
- [66] M. Palmer, O. Babko-malaya, A. Bies, M. Diab, M. Maamouri, A. Mansouri, and W. Zaghouni, ‘A pilot arabic propbank’, *Proceedings of the 6th International Language Resources and Evaluation, LREC*, vol. 8, pp. 28–30, 2008.
- [67] A. M. Al-Zoghby, A. S. E. Ahmed, and T. T. Hamza, ‘Arabic semantic web applications - A survey’, *Journal of Emerging Technologies in Web Intelligence*, vol. 5, no. 1, pp. 52–69, 2013.
- [68] S. T. Hashem A. J. Shmaisani, ‘USING SEMANTIC APPROACHES TO ANSWER ARABIC QUESTIONS FROM THE HOLY QURAN By Hashem A . J . Shmaisani Dr . Samir Tartir’, Philadelphia University, 2014.
- [69] S. Kulick, A. Bies, and M. Maamouri, ‘Consistent and Flexible Integration of Morphological Annotation in the Arabic Treebank’, no. Lrec, pp. 1–8, 2010.
- [70] F. Belkredim and A. El Sebai, ‘An Ontology Based Formalism for the Arabic Language Using Verbs and their Derivatives’, *Communications of the IBIMA*, vol. 11, pp. 44–52, 2009.
- [71] A. Farghaly and K. Shaalan, ‘Arabic Natural Language Processing: Challenges and Solutions’, *ACM Transactions on Asian Language Information Processing*, vol. 8, no. 4, pp. 1–22, 2009.

- [72] O. G. Alobaidi, K. a. Crockett, J. D. O’Shea, and T. M. Jarad, ‘Abdullah: An intelligent arabic conversational tutoring system for modern islamic education’, *Lecture Notes in Engineering and Computer Science*, vol. 2 LNECS, pp. 762–768, 2013.
- [73] H. Hasanuzzaman, ‘Arabic Language: Characteristics and Importance’, *ଅତ୍ୟନ୍ତର the Echo A journal of Humanities & Social Science*, vol. 1, no. 3, pp. 11–16, 2013.
- [74] K. Daimi, ‘Identifying Syntactic Ambiguities in Single-Parse Arabic Sentence’, no. 1989, pp. 333–349, 2001.
- [75] Y. Alotaibi and A. Meftah, ‘Review of distinctive phonetic features and the Arabic share in related modern research’, *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 21, no. 5, pp. 1426–1439, 2013.
- [76] S. Haboubi, S. Maddouri, and H. Amiri, ‘Identification of arabic word from bilingual text using character features’, 2011.
- [77] M. Attia, ‘Handling Arabic Morphological and Syntactic Ambiguity within the LFG Framework with a View to Machine Translation’, *Linguistics*, 2008.
- [78] M.-C. Landau, ‘Solving ambiguities in the semantic representation of texts’, *Proceedings of the 13th International Conference on Computational Linguistics*, pp. 239–244, 1990.
- [79] K. Shaalan, A. Farouk, and A. Rafea, ‘Towards An Arabic Parser for Modern Scientific Text’, no. April 2016, 1993.
- [80] S. Green and C. Manning, ‘Better Arabic parsing: Baselines, evaluations, and analysis’, *COLING ’10 Proceedings of the 23rd International Conference on Computational Linguistics*, no. August, pp. 394–402, 2010.
- [81] M. Z. Zemleni, R. Renken, J. C. J. Hoeks, J. M. Hoogduin, and L. A. Stowe, ‘Semantic ambiguity processing in sentence context: Evidence from event-related fMRI’, *NeuroImage*, vol. 34, no. 3, pp. 1270–1279, 2007.
- [82] J. M. Rodd, M. G. Gaskell, and W. D. Marslen-Wilson, ‘Modelling the effects of semantic ambiguity in word recognition’, *Cognitive Science*, vol. 28, no. 1, pp. 89–104, 2004.
- [83] J. M. Rodd, ‘The advantages and disadvantages of Semantic Ambiguity’, *Suite 101*, 2013.
- [84] H. Ishkewy, H. Harb, and H. Farahat, ‘Azharay: An Arabic Lexical Ontology’, *International journal of Web & Semantic Technology*, vol. 5, no. 4, pp. 71–82, 2014.
- [85] M. Attia, ‘An Ambiguity-Controlled Morphological Analyzer for Modern Standard Arabic Modelling Finite State Networks’, *Challenges of Arabic for NLP/MT Conference, The ...*, pp. 48–67, 2006.
- [86] E. Othman, K. Shaalan, and A. Rafea, ‘Towards resolving ambiguity in understanding Arabic sentence’, *International Conference on Arabic Language Resources and Tools*, p. 118–122, 2004.
- [87] S. K. Dwivedi and V. Singh, ‘Research and Reviews in Question Answering

- System', *Procedia Technology*, vol. 10, pp. 417–424, 2013.
- [88] N. A. Heba Kurdi, Sara Alkhaider and Department, 'DEVELOPMENT AND EVALUATION OF A WEB BASED QUESTION ANSWERING SYSTEM FOR ARABIC LANGUAGE', in *Logic-based approach for improving Arabic question answering. Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on. IEEE, 2014.* 2014, no. June, pp. 187–202.
- [89] M. Al-Yahya and H. Al-Khalifa, 'An Ontological Model for Representing Semantic Lexicons: An Application on Time Nouns in the Holy Quran', *The Arabian Journal for Science and Engineering*, vol. 35, no. 2, pp. 21–35, 2010.
- [90] E. Atwell, C. Brierley, K. Dukes, M. Sawalha, and A. Sharaf, '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*, pp. 1–13, 2011.
- [91] Y. Benajiba and P. Rosso, 'Arabic Named Entity Recognition using Conditional Random Fields', *Proc of Workshop on HLT & NLP within the Arabic World, LREC Vol 8*, pp. 143–153, 2008.
- [92] G. Kanaan, 'Building an Effective Rule-Based Light Stemmer for Arabic Language to Improve Search Effectiveness Riyadh Al-Shalabi Alaa Al-Nobani', *Science*, vol. 9, no. 4, pp. 368–372, 2008.
- [93] E. Atwell and A. Hardie, 'Second Workshop on Arabic Corpus Linguistics', no. July, pp. 0–10, 2013.
- [94] A. O. Al-Thubaity, 'A 700M+ Arabic corpus: KACST Arabic corpus design and construction', *Language Resources and Evaluation*, 2014.
- [95] E. Mohamed and S. Kübler, 'Arabic Part of Speech Tagging.', *Lrec*, pp. 2537–2543, 2010.
- [96] R. Mihalcea, S. Hassan, and E. Fernandez-ordo, 'Unsupervised Word Sense Disambiguation with Multilingual Representations', pp. 847–851, 2002.
- [97] L. Abouenour, K. Bouzoubaa, and P. Rosso, 'On the evaluation and improvement of Arabic WordNet coverage and usability', *Language Resources and Evaluation*, vol. 47, no. 3, pp. 891–917, 2013.
- [98] S. Elkateb, W. Black, S. St, P. Vossen, I. Technologies, D. Farwell, J. Girona, A. Pease, M. View, and C. Fellbaum, 'Arabic WordNet and the Challenges of Arabic', *The Challenge of Arabic for NLPMT International conference at the British Computer Society London 23 October 2006 pp1524 PDF 295KB*, no. Tufis 2004, pp. 15–24, 2006.
- [99] H. Rodríguez, D. Farwell, J. Farreres, M. Bertran, M. Alkhalifa, M. A. Martí, W. Black, S. Elkateb, J. Kirk, A. Pease, P. Vossen, and C. Fellbaum, 'Arabic wordnet: Current state and future extensions', *Proceedings of The Fourth Global WordNet Conference*, pp. 1–21, 2008.
- [100] E. Othman, K. Shaalan, and A. Rafea, 'A chart parser for analyzing modern standard Arabic sentence', *Proceedings of the MT Summit IX Workshop on Machine Translation for Semitic Languages*, pp. 37–44, 2003.

- [101] M. Maamouri, A. Bies, T. Buckwalter, and W. Mekki, 'The penn arabic treebank: Building a large-scale annotated arabic corpus', *NEMLAR Conference on Arabic Language Resources and Tools*, no. April 2016, pp. 102–109, 2004.
- [102] K. Dukes, E. Atwell, and A. M. Sharaf, 'Syntactic Annotation Guidelines for the Quranic Arabic Dependency Treebank', *In Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt*, pp. 1822–1827, 2010.
- [103] K. Dukes and N. Habash, 'Morphological Annotation of Quranic Arabic.', *Lrec*, pp. 1–3, 2010.
- [104] A. T. Al-taani and A. M. Al-gharaibeh, 'Searching Concepts and Keywords in the Holy Quran', *System*, pp. 1–3.
- [105] M. Almaayah, M. Sawalha, and M. A. M. Abushariah, 'A Proposed Model for Quranic Arabic WordNet', *Lre-Rel*, pp. 9–13, 2014.
- [106] B. Hammo and M. El-haj, 'Enhancing Retrieval Effectiveness of Diacritized Arabic Passages Using Stemmer and Thesaurus', *Artificial Intelligence*, no. APRIL, 2008.
- [107] J. Chandrika, B. Ramesh, K. R. Ananda, and R. D. Cunha, 'G Enetic a Lgorithm B Ased H Ybrid a Pproach for C Lustering T Ime', vol. 1, no. 1, pp. 39–52, 2014.
- [108] H. Hamam, M. T. Ben Othman, A. Kilani, F. Ncibi, and M. Ben Ammar, 'Exploring Qur ' an by using Aspects and Dependencies', vol. 5, no. 4, pp. 133–141, 2015.
- [109] ع.ع. ا. العلان، 'دور القرآن الكريم في حفظ معاني ألفاظ اللغة العربية في أصالتها وتطورها'، *دراسات علوم الشريعة والقانون*, vol. 41, no. 2, pp. 1191–1207, 2014.
- [110] 'خ. ا. خ. الكوسوفي، 'فضل القرآن الكريم ودوره في حفظ اللغة العربية وإثرائها *Majallah Al-qism Al-arabi*, no. 19, 2012.
- [111] S. Mohamed, A. Hassanin, and M. Othman, 'Educational System for the Holy Quran and Its Sciences for Blind and Handicapped People based on Google Speech API', *NooritcOrg*, no. March, pp. 150–161, 2014.
- [112] لإنجاز موسوعة إلكترونية شاملة للقرآن الكريم (prototype) محمد بن جماعة، 'نموذج تطبيقي مؤتمر جامعة طيبة الدولي في توظيف تقنية المعلومات لخدمة القرآن الكريم وعلومه، in 'وعلومه 2013, pp. 2-101-141.
- [113] A. S. Eljīl, 'The Exaggeration Formula of "Fuaa'la" –"to do" A Morphological and Semantic Study', *Journal Of Arabic And Human Sciences*, vol. 7, no. 2, 2014.
- [114] H. M. Al Chalabi, S. K. Ray, and K. Shaalan, 'Question classification for Arabic Question Answering Systems', *2015 International Conference on Information and Communication Technology Research, ICTRC 2015*, pp. 310–313, 2015.
- [115] M. Alqahtani and E. Atwell, 'Evaluation Criteria for Computational Quran Search', *IJA/SATSAT*, vol. 5, no. 1, pp. 12–22, 2017.

8 Appendices

8.1 Some examples of the result of applying corpus in question answering systems

الرقم	رقم السؤال	السؤال	2
2	لمن الحمد؟	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ	<p>الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (2)</p> <p>تِلْكَ آيَاتُ اللَّهِ نَتْلُوهَا عَلَيْكَ بِالْحَقِّ وَإِنَّكَ لَمِنَ الْمُرْسَلِينَ (252)</p> <p>وَمَنْ يَرْغَبْ عَنْ مِلَّةِ إِبْرَاهِيمَ إِلَّا مَنْ سَفِهَ نَفْسَهُ وَلَقَدْ اصْطَفَيْنَاهُ فِي الدُّنْيَا وَإِنَّهُ فِي الْآخِرَةِ لَمِنَ الصَّالِحِينَ (130)</p> <p>لَيْسَ عَلَيْكُمْ جُنَاحٌ أَنْ تَبْتَغُوا فَضْلًا مِنْ رَبِّكُمْ فَإِذَا أَفَضْتُمْ مِنْ عَرَفَاتٍ فَاذْكُرُوا اللَّهَ عِنْدَ الْمَشْعَرِ الْحَرَامِ وَاذْكُرُوهُ كَمَا هَذَاكُمْ وَإِنْ كُنْتُمْ مِنْ قَبْلِهِ لَمَنِ الضَّالِّينَ (198)</p> <p>وَاذْكُرُوا اللَّهَ فِي أَيَّامٍ مَعْدُودَاتٍ فَمَنْ تَعَجَّلَ فِي يَوْمَيْنِ فَلَا إِثْمَ عَلَيْهِ وَمَنْ تَأَخَّرَ فَلَا إِثْمَ عَلَيْهِ لِمَنِ اتَّقَى وَاتَّقُوا اللَّهَ وَاعْلَمُوا أَنَّكُمْ إِلَيْهِ تُحْشَرُونَ (203)</p> <p>مَثَلُ الَّذِينَ يُبْذِرُونَ أَمْوَالَهُمْ فِي سَبِيلِ اللَّهِ كَمَثَلِ حَبَّةٍ أَنْتَبَتْ سَبْعَ سَنَابِلٍ فِي كُلِّ سُنْبُلَةٍ مِائَةٌ حَبَّةٌ وَاللَّهُ يُضَاعِفُ لِمَنْ يَشَاءُ وَاللَّهُ وَاسِعٌ عَلِيمٌ (261)</p> <p>وَلَمَنْ أَنْتَبَتْ الَّذِينَ أُوْتُوا الْكِتَابَ بِكُلِّ آيَةٍ مَا تَبِعُوا قِبْلَتَكَ وَمَا أَنْتَ بِتَابِعٍ قِبْلَتَهُمْ وَمَا بَعْضُهُمْ بِتَابِعٍ قِبْلَةَ بَعْضٍ وَلَمَنِ اتَّبَعْتِ أَهْوَاءَهُمْ مِنْ بَعْدِ مَا جَاءَكَ مِنَ الْعِلْمِ إِنَّكَ إِذَا لَمِنَ الظَّالِمِينَ (145)</p> <p>بِاللَّهِ مَا فِي السَّمَوَاتِ وَمَا فِي الْأَرْضِ وَإِنْ تُبْذُوا مَا فِي أَنْفُسِكُمْ أَوْ تُخْفَوُهَا يُحَاسِبِكُمْ بِهِ اللَّهُ فَيَعْفُرُ لِمَنْ يَشَاءُ وَيُعَذِّبُ مَنْ يَشَاءُ وَاللَّهُ عَلَى كُلِّ شَيْءٍ قَدِيرٌ (284)</p> <p>وَاتَّبِعُوا مَا تَتْلُوا الشَّيَاطِينُ عَلَى مُلْكٍ سَلِيمٍ وَمَا كَفَرَ سَلِيمًا وَلَكِنَّ الشَّيَاطِينَ كَفَرُوا يُعَلِّمُونَ النَّاسَ السَّحَرَ</p>

<p>وَمَا أَنْزَلَ عَلَى الْمَلَائِكَةِ بَيِّنَاتٍ هَارُوتَ وَمَارُوتَ وَمَا يُعَلِّمَانِ مِنْ أَحَدٍ حَتَّى يَقُولَا إِنَّمَا نَحْنُ فِتْنَةٌ فَلَا تَكْفُرْ فَيَتَعَلَّمُونَ مِنْهُمَا مَا يُفَرِّقُونَ بِهِ بَيْنَ الْمَرْءِ وَزَوْجِهِ وَمَا هُمْ بِضَارِّينَ بِهِ مِنْ أَحَدٍ إِلَّا بِإِذْنِ اللَّهِ وَيَتَعَلَّمُونَ مَا يَضُرُّهُمْ وَلَا يَنْفَعُهُمْ وَلَقَدْ عَلِمُوا لَمَنِ اشْتَرَاهُ مَا لَهُ فِي الْآخِرَةِ مِنْ خَلَاقٍ وَلَبِئْسَ مَا شَرَوْا بِهِ أَنْفُسَهُمْ لَوْ كَانُوا يَعْلَمُونَ (102)</p> <p>وَأَتِمُّوا الْحَجَّ وَالْعُمْرَةَ لِلَّهِ فَإِنْ أُخْصِرْتُمْ فَمَا اسْتَيْسَرَ مِنَ الْهَدْيِ وَلَا تَخْلِقُوا رُءُوسَكُمْ حَتَّى يَبْلُغَ الْهَدْيُ مَحَلَّهُ فَمَنْ كَانَ مِنْكُمْ مَرِيضًا أَوْ بِهِ أَذًى مِنْ رَأْسِهِ فَقَدِّبْهُ مِنْ صِبْيَانٍ أَوْ صَدَقَةَ أَوْ نُسْلِكَ فَإِذَا أَمِنْتُمْ فَمَنْ تَمَنَّعَ بِالْعُمْرَةِ إِلَى الْحَجِّ فَمَا اسْتَيْسَرَ مِنَ الْهَدْيِ فَمَنْ لَمْ يَجِدْ فَصِيَامَ ثَلَاثَةِ أَيَّامٍ فِي الْحَجِّ وَسَبْعَةٍ إِذَا رَجَعْتُمْ تِلْكَ عَشْرَةٌ كَامِلَةٌ ذَلِكَ لِمَنْ لَمْ يَكُنْ أَهْلَهُ حَاضِرِي الْمَسْجِدِ الْحَرَامِ وَاتَّقُوا اللَّهَ وَاعْلَمُوا أَنَّ اللَّهَ شَدِيدُ الْعِقَابِ (196)</p>		
<p>يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَذَرُوا مَا بَقِيَ مِنَ الرِّبَا إِن كُنْتُمْ مُؤْمِنِينَ (278)</p> <p>نِسَاؤُكُمْ حَرْثٌ لَكُمْ فَأْتُوا حَرْثَكُمْ أَنَّى شِئْتُمْ وَقَدِّمُوا لِأَنْفُسِكُمْ وَاتَّقُوا اللَّهَ وَاعْلَمُوا أَنَّكُمْ مَلَاقُوهُ وَبَشِّرِ الْمُؤْمِنِينَ (223)</p> <p>وَلَا تَنْكِحُوا الْمُشْرِكَاتِ حَتَّى يُؤْمِنَ وَلَأَمَةٌ مُؤْمِنَةٌ خَيْرٌ مِنْ مُشْرِكَةٍ وَلَوْ أَعْبَبْتُمْ وَلَا تَنْكِحُوا الْمُشْرِكِينَ حَتَّى يُؤْمِنُوا وَلَعَبْدٌ مُؤْمِنٌ خَيْرٌ مِنْ مُشْرِكٍ وَلَوْ أَعْبَبَكُمْ أَوْلِيكُمُ الْجَنَّةُ وَالْمَغْفِرَةُ بِإِذْنِهِ وَيُبَيِّنُ آيَاتِهِ لِلنَّاسِ لَعَلَّهُمْ يَتَذَكَّرُونَ (221)</p> <p>الشَّهْرُ الْحَرَامُ بِالشَّهْرِ الْحَرَامِ وَالْحُرُمَاتِ قِصَاصٌ فَمَنْ اعْتَدَى عَلَيْكُمْ فَاعْتَدُوا عَلَيْهِ بِمِثْلِ مَا اعْتَدَى عَلَيْكُمْ وَاتَّقُوا اللَّهَ وَاعْلَمُوا أَنَّ اللَّهَ مَعَ الْمُتَّقِينَ</p> <p>نِسَاؤُكُمْ حَرْثٌ لَكُمْ فَأْتُوا حَرْثَكُمْ أَنَّى شِئْتُمْ وَقَدِّمُوا لِأَنْفُسِكُمْ وَاتَّقُوا اللَّهَ وَاعْلَمُوا أَنَّكُمْ مَلَاقُوهُ وَبَشِّرِ الْمُؤْمِنِينَ</p> <p>الشَّهْرُ الْحَرَامُ بِالشَّهْرِ الْحَرَامِ وَالْحُرُمَاتِ قِصَاصٌ فَمَنْ اعْتَدَى عَلَيْكُمْ فَاعْتَدُوا عَلَيْهِ بِمِثْلِ مَا اعْتَدَى عَلَيْكُمْ</p>	<p>كُنِبَ عَلَيْكُمْ إِذَا حَصَرَ أَحَدُكُمْ الْمَوْتَ إِنْ تَرَكَ خَيْرًا الْوَصِيَّةَ لِلْأَقْرَبِينَ بِالمَعْرُوفِ حَقًّا عَلَى الْمُتَّقِينَ</p> <p>الشَّهْرُ الْحَرَامُ بِالشَّهْرِ الْحَرَامِ وَالْحُرُمَاتِ قِصَاصٌ فَمَنْ اعْتَدَى عَلَيْكُمْ فَاعْتَدُوا عَلَيْهِ بِمِثْلِ مَا اعْتَدَى عَلَيْكُمْ وَاتَّقُوا اللَّهَ وَاعْلَمُوا أَنَّ اللَّهَ مَعَ الْمُتَّقِينَ</p> <p>نِسَاؤُكُمْ حَرْثٌ لَكُمْ فَأْتُوا حَرْثَكُمْ أَنَّى شِئْتُمْ وَقَدِّمُوا لِأَنْفُسِكُمْ وَاتَّقُوا اللَّهَ وَاعْلَمُوا أَنَّكُمْ مَلَاقُوهُ وَبَشِّرِ الْمُؤْمِنِينَ</p>	<p>12</p> <p>بم وصف الله تعالى المتقين المؤمنين</p>

<p>وَاتَّقُوا اللَّهَ وَعَلِمُوا أَنَّ اللَّهَ مَعَ الْمُتَّقِينَ (194)</p> <p>أَمَّنَ الرَّسُولُ بِمَا أُنزِلَ إِلَيْهِ مِنْ رَبِّهِ وَالْمُؤْمِنُونَ كُلٌّ آمَنَ بِاللَّهِ وَمَلَائِكَتِهِ وَكُتُبِهِ وَرُسُلِهِ لَا نُفَرِّقُ بَيْنَ أَحَدٍ مِنْ رُسُلِهِ وَقَالُوا سَمِعْنَا وَأَطَعْنَا غُفْرَانَكَ رَبَّنَا وَإِلَيْكَ الْمَصِيرُ (285)</p> <p>وَلِلْمُطَلَّاتِ مَتَاعٌ بِالْمَعْرُوفِ حَقًّا عَلَى الْمُتَّقِينَ (241)</p> <p>وَإِذَا قِيلَ لَهُمْ آمِنُوا بِمَا أَنْزَلَ اللَّهُ قَالُوا نُؤْمِنُ بِمَا أُنزِلَ عَلَيْنَا وَيَكْفُرُونَ بِمَا وَرَاءَهُ وَهُوَ الْحَقُّ مُصَدِّقًا لِمَا مَعَهُمْ قُلْ فَلِمَ تَقْتُلُونَ أَنْبِيَاءَ اللَّهِ مِنْ قَبْلِ أَنْ كُنْتُمْ مُؤْمِنِينَ (91)</p> <p>ذَلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ (2)</p> <p>فَجَعَلْنَاهَا نَكَالًا لِّمَا بَيْنَ يَدَيْهَا وَمَا خَلْفَهَا وَمَوْعِظَةً لِّلْمُتَّقِينَ (66)</p> <p>لَيْسَ الْبِرَّ أَنْ تُولُوا وَجُوهَكُمْ قِبَلَ الْمَشْرِقِ وَالْمَغْرِبِ وَلَكِنَّ الْبِرَّ مَنْ آمَنَ بِاللَّهِ وَالْيَوْمِ الْآخِرِ وَالْمَلَائِكَةِ وَالْكِتَابِ وَالنَّبِيِّينَ وَآتَى الْمَالَ عَلَى حُبِّهِ ذَوِي الْقُرْبَى وَالْيَتَامَى وَالْمَسَاكِينَ وَابْنَ السَّبِيلِ وَالسَّائِلِينَ وَفِي الرِّقَابِ وَأَقَامَ الصَّلَاةَ وَآتَى الزَّكَاةَ وَالْمُوفُونَ بِعَهْدِهِمْ إِذَا عَاهَدُوا وَالصَّابِرِينَ فِي الْبَأْسَاءِ وَالضَّرَّاءِ وَحِينَ الْبَأْسِ أُولَئِكَ الَّذِينَ صَدَقُوا وَأُولَئِكَ هُمُ الْمُتَّقُونَ (177)</p>			
<p>يَمْحَقُ اللَّهُ الرَّبَّاءِ وَيُرِيهِ الصَّدَقَاتِ وَاللَّهُ لَا يُحِبُّ كُلَّ كَفَّارٍ أَثِيمٍ (276)</p> <p>إِنَّ الَّذِينَ كَفَرُوا وَمَاتُوا وَهُمْ كُفَّارٌ أُولَئِكَ عَلَيْهِمْ لَعْنَةُ اللَّهِ وَالْمَلَائِكَةِ وَالنَّاسِ أَجْمَعِينَ (161)</p> <p>بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ (1)</p> <p>وَقَالُوا اتَّخَذَ اللَّهُ وَلَدًا سُبْحَانَ اللَّهِ بَلْ لَمْ يَكُنْ فِي السَّمَوَاتِ وَالْأَرْضِ كُلُّ لَهٍ قَانِتُونَ</p>	0	<p>بم وصف الله تعالى الكفار والمناققين</p>	13

<p>(116) صِبْغَةَ اللَّهِ وَمَنْ أَحْسَنُ مِنَ اللَّهِ صِبْغَةً وَنَحْنُ لَهُ عَابِدُونَ (138)</p> <p>إِنَّ الَّذِينَ آمَنُوا وَالَّذِينَ هَاجَرُوا وَجَاهَدُوا فِي سَبِيلِ اللَّهِ أُولَئِكَ يَرْجُونَ رَحْمَةَ اللَّهِ وَاللَّهُ غَفُورٌ رَحِيمٌ (218)</p> <p>مَنْ ذَا الَّذِي يُفْرِضُ اللَّهُ قَرْضًا حَسَنًا فَيُضَاعِفُهُ لَهُ أَضْعَافًا كَثِيرَةً وَاللَّهُ يَقْبِضُ وَيَبْسُطُ وَإِلَيْهِ تُرْجَعُونَ (245)</p> <p>فَإِنْ انْتَهَوْا فَإِنَّ اللَّهَ غَفُورٌ رَحِيمٌ (192)</p> <p>فَهَزَمُوهُمْ بِإِذْنِ اللَّهِ وَقَتَلَ دَاوُدُ جَالُوتَ وَآتَاهُ اللَّهُ الْمُلْكَ وَالْحِكْمَةَ وَعَلَّمَهُ مِمَّا يَشَاءُ وَلَوْلَا دَفْعُ اللَّهِ النَّاسَ بَعْضَهُمْ بِبَعْضٍ لَفَسَدَتِ الْأَرْضُ وَلَكِنَّ اللَّهَ ذُو فَضْلٍ عَلَى الْعَالَمِينَ (251) (وَبِاللَّهِ الْمَشْرِقِ وَالْمَغْرِبِ فَأَيْنَمَا تُوَلُّوا فَتَمَّ وَجْهَ اللَّهِ إِنَّ اللَّهَ وَاسِعٌ عَلِيمٌ (115))</p>			
<p>يُخَادِعُونَ اللَّهَ وَالَّذِينَ آمَنُوا وَمَا يَخْدَعُونَ إِلَّا أَنفُسَهُمْ وَمَا يَشْعُرُونَ (9)</p> <p>160</p>	<p>يُخَادِعُونَ اللَّهَ وَالَّذِينَ آمَنُوا وَمَا يَخْدَعُونَ إِلَّا أَنفُسَهُمْ وَمَا يَشْعُرُونَ</p> <p>1</p>	<p>هل يشعرون الذين يخدعون الله بذلك</p>	<p>16</p>
<p>السؤال بعد التعديل هو : هل يستهزئ له كافر</p> <p>total matching 165 documents</p> <p>اللَّهُ يَسْتَهْزِئُ بِهِمْ وَيَمُدُّهُمْ فِي طُغْيَانِهِمْ يَعْمَهُونَ (15)</p>	<p>اللَّهُ يَسْتَهْزِئُ بِهِمْ وَيَمُدُّهُمْ فِي طُغْيَانِهِمْ يَعْمَهُونَ</p>	<p>هل يستهزئ الله بالكافرين</p>	<p>22</p>
<p>السؤال بعد التعديل هو : اعدت نار قود</p> <p>total matching 14 documents</p> <p>فَإِنْ لَمْ تَفْعَلُوا وَلَنْ تَفْعَلُوا فَاتَّقُوا النَّارَ الَّتِي وَقُودُهَا النَّاسُ وَالْحِجَارَةُ أُعِدَّتْ لِلْكَافِرِينَ وَالَّذِينَ كَفَرُوا وَكَذَّبُوا بِآيَاتِنَا أُولَئِكَ أَصْحَابُ النَّارِ هُمْ فِيهَا خَالِدُونَ</p> <p>فَإِنْ لَمْ تَفْعَلُوا وَلَنْ تَفْعَلُوا فَاتَّقُوا النَّارَ الَّتِي وَقُودُهَا النَّاسُ وَالْحِجَارَةُ أُعِدَّتْ</p>	<p>فَإِنْ لَمْ تَفْعَلُوا وَلَنْ تَفْعَلُوا فَاتَّقُوا النَّارَ الَّتِي وَقُودُهَا النَّاسُ وَالْحِجَارَةُ أُعِدَّتْ لِلْكَافِرِينَ وَالَّذِينَ كَفَرُوا وَكَذَّبُوا بِآيَاتِنَا أُولَئِكَ أَصْحَابُ النَّارِ هُمْ فِيهَا خَالِدُونَ</p> <p>وَقَالُوا لَنْ نَمَسَّنَا النَّارَ إِلَّا أَيَّامًا مَعْدُودَةً قُلْ أَتَّخَذْتُمْ عِنْدَ اللَّهِ عَهْدًا فَلَنْ يُخْلَفَ اللَّهُ عَهْدَهُ أَمْ تَقُولُونَ</p>	<p>لم اعدت النار وما هو وقودها</p>	<p>31</p>

<p>لِلْكَافِرِينَ (24)</p>	<p>عَلَى اللَّهِ مَا لَا تَعْلَمُونَ</p> <p>بَلَىٰ مَنْ كَسَبَ سَيِّئَةً وَأَحَاطَتْ بِهِ خَطِيئَتُهُ فَأُولَٰئِكَ أَصْحَابُ النَّارِ هُمْ فِيهَا خَالِدُونَ</p> <p>وَإِذْ قَالَ إِبْرَاهِيمُ رَبِّ اجْعَلْ هَذَا بَلَدًا آمِنًا وَارْزُقْ أَهْلَهُ مِنَ الثَّمَرَاتِ مَنْ آمَنَ مِنْهُمْ بِاللَّهِ وَالْيَوْمِ الْآخِرِ قَالَ وَمَنْ كَفَرَ فَأُمَتِّعُهُ قَلِيلًا ثُمَّ أَضْطَرُّهُ إِلَى عَذَابِ النَّارِ وَبِئْسَ الْمَصِيرُ</p> <p>إِنَّ الَّذِينَ يَكْتُمُونَ مَا أَنْزَلَ اللَّهُ مِنَ الْكِتَابِ وَيَسْتَتِرُونَ بِهِ تَمَنَّا قَلِيلًا أُولَٰئِكَ مَا يَأْكُلُونَ فِي بُطُونِهِمْ إِلَّا النَّارَ وَلَا يَكَلِّمُهُمُ اللَّهُ يَوْمَ الْقِيَامَةِ وَلَا يُزَكِّيهِمْ وَلَهُمْ عَذَابٌ أَلِيمٌ</p> <p>أُولَٰئِكَ الَّذِينَ اسْتَنَرُوا الضَّلَالَةَ بِالْهَدَىٰ وَالْعَذَابِ بِالْمَغْفِرَةِ فَمَا أَصْبَرَهُمْ عَلَى النَّارِ</p> <p>اللَّهُ وَلِيُّ الَّذِينَ آمَنُوا يُخْرِجُهُم مِنَ الظُّلُمَاتِ إِلَى النُّورِ وَالَّذِينَ كَفَرُوا أُولَٰئِكَ هُمُ الطَّاغُوتُ يُخْرِجُونَهُمْ مِنَ النُّورِ إِلَى الظُّلُمَاتِ أُولَٰئِكَ أَصْحَابُ النَّارِ هُمْ فِيهَا خَالِدُونَ</p>	
----------------------------	---	--