USING SEMANTIC APPROACHES TO ANSWER ARABIC QUESTIONS FROM THE HOLY QURAN

By

Hashem A. J. Shmaisani

Supervisor
Dr. Samir Tartir

This Thesis was Submitted in Partial Fulfilment of the Requirements for the Master's Degree in Computer Science

Deanship of Academic Research and Graduate Studies
Philadelphia University

2014

جامعة فيلادلفيا
نموذج تفويض

أنا هاشم عدنان جدعان شميساني، أفوض جامعة فيلادلفيا بتزويد نسخ من رسالتي للمكتبات أو المؤسسات أو الهيئات أو الأشخاص عند طلبها.

التوقيع :
التاريخ :

Philadelphia University
Authorization Form

I am, Hashem A. J. Shmaisani, authorize Philadelphia University to supply copies of my thesis to libraries or establishments or individuals upon request.

Signature:

Date:

USING SEMANTIC APPROACHES TO ANSWER ARABIC QUESTIONS FROM THE HOLY QURAN

By

Hashem A. J. Shmaisani

Supervisor
Dr. Samir Tartir

This Thesis was Submitted in Partial Fulfilment of the Requirements for the Master's Degree in Computer Science

Deanship of Academic Research and Graduate Studies
Philadelphia University

2014

Successfully defended and approved on _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

Examination Committee Signature		**Signature**
Dr. Academic Rank:	, Chairman.	_ _ _ _ _ _ _ _ _ _ _ _
Dr. Academic Rank:	, Member.	_ _ _ _ _ _ _ _ _ _ _ _
Dr. Academic Rank:	, Member.	_ _ _ _ _ _ _ _ _ _ _ _
Dr. Academic Rank:	, External Member.	_ _ _ _ _ _ _ _ _ _ _ _

Dedication

First of all I thank Allah the almighty for giving me the strength and knowledge to finish this work, I dedicate this work to my family and my mother especially, she was there for me whenever I needed, last but not least I dedicate this work also to my supervisor *Dr. Samir Tartir* who was helping me unconditionally, he was like a big brother to me rather than an academic supervisor.

Hashem A. J. Shmaisani
2014

Acknowledgment

(قَالُواْ سُبْحَانَكَ لاَ عِلْمَ لَنَا إِلاّ مَا عَلّمْتَنَا إِنّكَ أَنتَ الْعَلِيمُ الْحَكِيمُ)

It would not have been possible to write this master thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

Above all, I would like to express my thanks and sincere gratitude for who has guided me through my study and my thesis work; my supervisor *Dr. Samir Tartir*, for giving me strength, support and knowledge in exploring things, also I was honoured to benefit from the comments and guidance of Prof. *Saeed Al-Ghoul*, Dr. *Nameer El-Emam* and Dr. *Majdi Sawalha*.

Finally, I would like to thank my parents for giving me their unequivocal support, and to those who supported and encouraged me in any way; my teachers, friends, superiors and colleagues at Philadelphia University.

Hashem A. J. Shmaisani
2014

Table of Contents

List of Appendices

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Full Name
QASAL	Question Answering system for the Arabic language
AQuASys	Arabic Question Answering system
QAS	Question Answering system
DefArabicQA	Arabic definition question answering
AQO	Arabic Quranic ontology
UNESCO	United Nations Organisation for Education, Science, Culture and Communications

Abstract

Arabic is the fifth language in the world according to number of speakers, and is considered a second language to over one billion Muslims worldwide. Yet, very little research has been done on machine processing of Arabic text.

One of the main methods that is effected by language in computer science research areas, is the searching capability quality, Arabic language search is still in the phase of syntax based search, retrieving the exact match to the user based on the provided search keywords syntax, only some tried to move beyond that, but still no work was done regarding searched data structure or semantic Question Answering methodologies to serve this language.

Relying on the fact that in general languages are semantic based more than syntax based, a single word may give more than one meaning based on the context, it holds more meaning than just a string of characters.

Accordingly, we are proposing an approach to overcome the semantic and structural challenges that are currently limiting the quality of search among Arabic based content. Mainly the goal is to solve semantic issues by using synonyms for the user query keywords, as for the content structure issues an ontology based approach is used in order to organize textual entities and their relationships with other content in the text, rather than simple strings of characters.

To test and implement this approach for the first time a very unique semantic based context is used, which is the Holy Quran. Nowadays people searching the Holy Quran for knowledge are using term and exact-match based approaches, which is not suitable for a rich text as the Quran. People searching Holy Quran are often looking for a concept among the rather than a certain text. In other words, they are looking for the semantic similarity rather than the syntactic similarity. Which explains the use of a semantic layer in our system. Furthermore, the type of queries coming from the Holy Quran searchers are question based queries [e.g. How many years did the people of the cave slept?], so in order to cover this

side a question analysis layer is embedded in order to detect keywords and understand users' questions in order to determine the desired answer.

The used ontology content is based on the relation extractor model, which is a very important contribution to the Arabic ontology building. The relation extracting process is the action of detecting and retrieving relations and their concepts among raw text. This thesis is proposing a model that can detect relations and concepts among any given Arabic context depending on grammatical perspectives. Which is, up to our knowledge, the first semantic relation extractor for the Arabic language.

In summary, this research builds a question answering layer over a semantic layer built over an ontology layer. This will be implemented for the first time for the Arabic language in general and for the Holy Quran in precise.

CHAPTER ONE: INTRODUCTION

1.1 Preface

This thesis is providing an approach to model Arabic language raw unstructured text into a semantic-relation based structure that aids efficient semantic search in terms that can handle question answering. In order to test this approach the Arabic text of the Holy Quran is selected to be the targeted text.

1.2 Research Context

1.3 The Arabic language

The Arabic language is one of the main spoken languages in the world it is used for human communication in a wide arc of territory stretching across the Middle East, North Africa, and the Horn of Africa, Arabic belongs to the Afro-Asiatic language family, it is spoken by as many as 422 million, first language speakers, making it one of the half dozen most populous languages in the world, and ranked fifth according to (UNESCO, 2012).

Although this information shows how widely used Arabic is, it is still far behind in works in the computer science community such as search engines, question answering systems and NLP algorithms. This research is presenting an approach that will change and enhance the way Arabic language is used and modelled in these domains.

1.4 The Holy Quran

The Holy Quran is the highest accepted source of religious legislation for Muslims since they believe it to be the word of Allah that was delivered to the angel Gabriel whom delivered it directly to the prophet Mohammed, this was nearly fourteen hundred years ago as revelations started in the year 609 ending in 632, (Ibn-katheer, 1370).

Due to this belief, the Holy Quran is the main source which 1.5 billion Muslims all around the world find their religious teachings, knowledge and rulings, ranging from inheritance to history and governance. The major problem that is currently facing anybody trying to find answers from the Holy Quran. Specifically, there is no Arabic Question Answering system that can handle their query semantically, as questions may not contain

the exact words mentioned in the Holy Quran, but synonyms of the searched words, or may even contain meaning (concept) of the word but using different terms (synonyms).

1.5 Problem Statement

Until this moment and up to our knowledge, there is no Arabic question answering system that people can use to find accurate answers to questions they have in the Holy Quran. In the past 4 years many approaches were attempted (Ezzeldin and Shaheen, 2012), but still answers were not accurate and most of the approaches rely on a very naïve exact match approach. Web based search tools for the Holy Quran give users the ability by which the user can get back all the verses where the exact syntax match for the terms he searched for appears, which is often impractical to most users, because when searching the Holy Quran the goal in most cases is the semantic relativeness to a certain concept, not only finding the exact match, other cases involve the user looking for an answer for his question, the reason why the Holy Quran was selected for our approach is because it is a very rich context with concepts and relations that is very hard to represent or search using the currently used syntactic based approaches no system currently do this functionality for the Holy Quran.

For example, a search using the word "مسيحي" (Christian) as a keyword will not return any results in most of the current tools, even though the Holy Quran talks about Christians and Christianity in many locations. The reason for this lack of answers is that the word used for Christian in the text of the Holy Quran is "نصارى" in the plural form. And due to the fact that current approaches use naïve syntax search the word is not found. While searching for the word "نصارى" should return 13 results in any quranic search engine.

Therefore, our approach attempts to utilize the Quranic syntax organizational structure and semantic aspect to be able to understand any user questions and find the most accurate answer.

Semantic question answering systems are systems that allow users to query semantic data (Lopez et al, 2005) using natural language questions (Tartir et al, 2009), and this is exactly what is actually intended when querying Quranic verses.

Furthermore correct annotation of the Holy Quran requires not only deep understanding of Arabic linguistics (Hammo et al, 2002), but also of the source material (Dukes, 2011), the Holy Quran itself. Given the importance of the Holy Quran to the Islamic faith, any syntactic annotation needs to be carefully considered since alternative parses for a sentence can suggest alternative meanings for the scripture in certain cases. Fortunately, the unique form of Arabic in which the Holy Quran has been inscribed in, has been studied in detail for over 1,000 years. This is far longer than corresponding grammars for most other languages, and in fact traditional Arabic grammar is considered to be one of the origins of modern dependency Grammar, (Atwell et al, 2011).

Nowadays searching is done in a strict keyword match based form, or in the case of some approaches that added a semantic layer that adds synonym alternatives to the searched keywords, but still, the search is working over unstructured data that is treated as simple strings of text that have no relation or related concepts whatsoever, whereas in fact the main feature of any smart search approach is using the semantic relational power that relies within linguistics in any given language, Semantic search seeks to improve search accuracy by understanding the searcher intent and the contextual meaning of terms as they appear in the searchable data space, part of semantics is using synonyms and defining the terms by the meaning not by the exact syntax match.

We are proposing a hypothesis that we can benefit from the relations found within any context in order to structure the content which can create a semantic based content rather than a purely syntactic string of characters that does not entail any meaning.

Using the previously proposed semantic structure we can aid semantic search in the terms of searching using relations that can be detected within the user query then can be used to detect related concepts in a context inspite of the fact that some of these

related concepts can be located in non-contiguous form, furthermore these semantic relation based structures can benefit question answering capabilities as in our question answering system, because when each concept is attached to its related concepts we will end up with the capability of answering any question about this concept.

1.6 Motivation

The main motivations that encouraged this research are the current limitations and insufficiencies within Arabic search engines; quranic search engines in particular, also the emerging need of new smarter search and question answering techniques for the Arabic language.

One of the main limitations was the absence of any semantic feature in most of the current online engines, for an example if a user wants to find a concept within a text and this concept wasn't mentioned as an exact match in this text but it was mentioned by its meaning, in this case the user won't be able to retrieve any results, because current Arabic search engines use exact match approaches, some better Arabic search engines give back root derived versions for the searched concept, but still this doesn't consider the importance of relations that rely within the searched concepts and it doesn't have question answering capabilities.

Relational connectors absence is another main limitations among the structure that holds the data or texts which will be searched, relational connectors are connectors that indicate the relation between concepts and their objects/individuals throughout a certain text, the main benefit from these connectors is that they connect all related data to a certain concept, this paves the road for the implementation of smart approaches on the data, such as question answering systems and search engines.

Question answering feature within a search engine is a hot research topic nowadays, the only way to answer precise answers is working within a database where we have a strict structure to place data and a closed world assumptions, which means that we have very specific types of question patterns to use.

Another main limitation is that nowadays users trying to find information in Arabic texts have to enter the exact syntax match for their desired string into search engines, this will only retrieve paragraph snippets that surround the strings entered by the user, so retrieved results are text snippets that contains these keywords, obviously this isn't practical to users looking for direct answers to certain questions, or even searching for a simple keyword but don't know the exact syntax match mentioned in the targeted text.

1.7 Contributions

In this research we are proposing a 3 layered framework that consists of using an ontology to structure data, a semantic layer to enhance semantics of the search process and a question answering layer to handle question based queries.

The main contributions in this research were building the first Arabic relation extractor, which was needed to build this extractor to aid the evolving research area of ontologies and semantic web, and this extractor uses the grammatical structure of the Arabic language to detect relations and concepts.

Another contribution is building the first Arabic ontology for the Holy Quran, this ontology will aid any future work in Quranic search approaches, and this structure shows and highlights the semantics that relies within the Holy Quran.

1.8 Thesis layout

This documentation is structured as following, in chapter two domain background and related work will be introduced along with some comparisons and summaries of their features, after that, this thesis working for approach will be introduced in details, in chapter four more implementation details will be introduced along with visual figures of the work, in chapter five we will evaluate and compare results with other systems, in the end, conclusion and future work guidelines will be introduced.

CHAPTER TWO: BACKGROUND / RELATED WORK

2.1 Introduction

In this chapter, we will introduce related work to each of our proposed layers in our question answering system, we will show strength aspects that we learned from and also we will show the insufficiencies in the previous works and show how we solved these issues in this research.

2.2 Quranic search techniques

Previous related work in the Quranic search techniques relied on commonly used naïve Keyword-In-Context search approaches to find exact examples of a word in the Quranic text, in order to detect the occurrences of that word, or even get the verses where that word appeared, these approaches mainly suffered from the absence of any semantic feature in their search structure, also they didn't have any question answering feature in them.

2.2.1 QURANY[1]

Qurany meaning ("my Quran" in Arabic) by Abbas, (2009), incorporates novel features to improve recall and precision when compared against other tools for searching for concepts in the Holy Quran. To increase recall, the system accepts search terms in English and/or Arabic, and the underlying corpus includes eight variant English translations as well as the Arabic original. So, a verse can be found if any of its translations includes the specific word form; more of the desired verses are found, improving the chances of finding all desired verses. The system also lemmatizes the keyword to improve recall further. However, keyword search is unlikely to be effective for more abstract concepts which Muslim and other scholars may want to examine, such as 'Juridical Regulations' or 'Women's Property'. This is because these are abstract concepts that are described in several different chapters and verses of the Holy Quran but not explicitly labelled as 'Juridical Regulations' or 'Women's Property'.

As for our proposed approach, it will be more semantic based, we will not use concepts only to answer a user query, we will utilize ontology relations among these concepts along

[1] www.Quranykeywords.appspot.com/

with the aid of the synonym generation layer, these features was not taken care of in Qurany.

2.2.2 Text mining the Quran[2]

This is an approach to the Quranic text from a data mining point of view presented by Sharaf, (2012), he assumed that the Holy Quran holds lots of interesting information, facts, correlations, patterns, associations between facts and concepts that are difficult to discover by manual naïve syntax processing.

Hence, he tried to employ computational techniques from the fields of text mining, machine learning, natural language processing, computational linguistics and stylometrics to reveal some of the hidden trends and make it easy to link the scattered yet related concepts in the Quran. The research on this project is still emerging getting aid from computational fields and respects the guidelines set by early Quranic scholars and their books of Tafseer like Ibn Jarir at-Tabari, Al-Baghawi and Ibn Kathir.

This research created a large heap toward adding intelligence to the Quranic aiding tools and software, in spite of that, still no question answering features are forced, no "full" ontology is used, and no semantic backup is used, in our proposed system we will cover these very important features.

2.2.3 Similarity based search for the Holy Quran

An approach to widen search results by using simple and naïve parsing techniques in order to search for syntax similarities for the searched key words, note that they produce syntactically similar keywords, not semantically similar, so this is not a semantic approach at all, this approach was introduced in some online Quranic search engines, yet no semantics nor ontologies was used at all, al "MOYSAR"[3] Quranic search engine is one of these examples.

---

[2] www.textminingtheQuran.com
[3] www.moysar.com/searchq.php

2.2.4 Quran corpus system[4]

Actually this is a very good approach that designed a Quranic online tool by Dukes, (2011), an Arabic language grammar based parsing for most of the main verses in the Holy Quran and a partial "not full" ontology for the Quranic text, nevertheless, no semantics are declared regarding meaning and tafseer of each verse, also no question answering approach is mentioned at all.

2.2.5 Arabic corpus online system[5]

One of the oldest and largest projects in this domain is the Arabic corpus, developed and maintained by D. Parkinson, this tool added a great feature of giving semantics for words users might search for, such as what part of speech is it, its meaning and other semantic services. Yet no ontology or question answering model was used to aid the search methodology in Arabic corpus, nevertheless Arabic corpus is the base for any new semantic research on the Arabic language.

2.2.6 The computational quranic linguistics team, university of Leeds[6]

One of the most distinguished works and techniques done for the Holy Quran was done by this group of researchers, located in Leeds University in England, their research was mainly on the Arabic language and on the Quranic field in particular, they have some interesting research in the field of developing researching techniques for the Holy Quran, here are some Computation Corpus linguistic research projects on the Quran:

- Qurany subject browser.
- Quranic Arabic corpus.
- Tools on text mining the Quran.

As for future works they proposed projects for future research which are:

- Quranic knowledge map project.
- Text mining hadith project.

---

[4] www.corpus.Quran.com/Qurandictionary.jsp
[5] www.arabicorpus.byu.edu
[6] http://www.comp.leeds.ac.uk/arabic

- Arabic Question answering systems

2.2.7 Comparison for the current Quranic search techniques

Table 1 - Comparison for the current Quranic search techniques

Criteria / Tool	Covered semantics	Ontology presence	Question Answering	Data mining techniques	Application released
QURANY	No	No	No	No	Yes
Quran mining	No	No	No	Yes	No
Quran similarity	No	No	No	No	Yes
Quran corpus	Yes	Semi – onto	No	No	Yes
Arabic corpus	Yes	No	No	No	Yes

2.3 Arabic question answering systems

Question answering systems is a very old yet still evolving research topic, early question answering systems relied on a keyword and question pattern recognition techniques to detect and analyse questions, also they relied on raw data that was searched among unstructured documents, then a ranking process is initiated to rank retrieved snippets of text according to keyword reoccurrence, more sophisticated question answering systems embedded a semantic level to enhance search results by retrieving document snippets that contain the searched concept by its meaning not by the exact match.

2.3.1 QARAB

QARAB is a question answering system developed by Hammo et al, (2002), to answer questions from Arabic documents, the approach of QARAB starts by processing the users question tokenizing it and removing stop words then run grammatical tagging for the keywords, these keywords are sent to a query that will be used on a database of documents where it will detect the weight for each keyword, then ranked snippets containing the answer is retrieved, finally an answer is generated based on the question pattern.

2.3.2 QASAL

QASAL by Brini et al, (2009), is a question answering system for the Arabic language, it is composed of three layers, the first layer is a question analyser that handles the user question in the terms of tokenizing and keyword extraction, the second layer is a passage retrieval layer that retrieved related passages to the targeted keywords and the third layer is an answer extraction layer that extracts the answer upon a predefined answer pattern.

2.3.3 AQuASys

AQuASys Arabic question answering system was developed by Bekhti and Al-Harbi,(2011), to answer unformatted questions by detecting keywords and answer pattern identifiers like the words (where, when, who…etc.), the addition in this system was that the researchers widen the pattern identifiers set to cover more questions and answer patterns.

2.3.4 QAS

QAS was developed by Bdour and Gharaibeh, (2013), as a question answering system that produce yes/no answers, QAS starts with parsing the question and removing stop words, tagging the keywords with grummer tags and removing negation words if any, the next step is retrieving text snippets that surround the desired keywords, after that the snippet is tested to fid negation words if any, if the sentence is negated then the answer is no, if no negation words was found then the answer is yes.

2.3.5 DefArabicQA

The Arabic definition question answering system was developed by Trigui et al, (2010), to answer questions but only definition question patterns, it works on web resources to retrieve candidate answer snippets, as a first step the question pattern is detected by analysing the question, then a predefined corresponding answer patterns are used to find answers from web resources.

2.3.6 Comparison of Arabic Question Answering systems

Table 2 - Comparison of Arabic Question Answering systems

Criteria / Tool	Covered semantics	Ontology presence	Restricted pattern forms	Data mining techniques	Application released

QARAB	No	No	No	No	No
QASAL	No	No	Yes	No	No
AQuASys	No	No	Yes - enhanced	No	No
QAS	No	No	Yes	No	No
DefArabicQA	No	No	Yes	Yes	No

2.4 English ontology based question answering systems

2.4.1 OntoQA

A very significant addition was presented by Tartir and Arpinar, (2013), they proposed the OntoQA system where a layer of ontology was added to semantically relate concepts in order to get better question answering results, in this system the first step is to calculates metric values that identify the importance of concepts within the predefined ontology, then Wordnet is used to add semantically related words to the search query, then candidate answers are generated from the ontology according to keyword weight, finally a list of answers is generated.

2.4.2 Ginsing

Ginseng was introduced by Bernstein et al, (2006), Ginsing is a system that uses NLP to aid users with suitable suggestions while writing their desired search question in order to preserve a suitable form to search the used ontology with. After detecting the user question Ginseng uses the detected question and transforms it into a query that runs against the ontology to retrieve the desired result.

2.4.3 AquaLog

AquaLog was introduced by Lopez et al, (2005), AquaLog was developed to shoe the question answering abilities that can be generated if using an ontology, AquaLog input is an ontology that a user wants to query and the questions that the user wants to get, results about from the ontology, simply the output is answers derived from the previously used ontology.

2.5 Arabic and Quran ontologies

Quran ontologies was a hot research topic in the past 3 years, most approaches tried to show the semantic strength of the Arabic language through the Holy Quran, and here are some of these works.

2.5.1 Quran corpus[7]

Quran ontology was introduced for the first time by Kais Dukes in (2009), where Dukes used a class hierarchy representation to define the key concepts in the Quran and shows relations between them, after that he worked on the grammatical dependencies in the Quran in order to create a grammatical dependency tree that can help researchers in the Quran ontology in the future.

2.5.2 Quran time

One of the first Arabic based ontologies was introduced by Al-Yahya et al, (2010), this was a research to represent lexicons related to time concepts in the Holy Quran, this was done by detecting all time related concepts in the Holy Quran and organizing them into an ontology in a semantic relational way, all related concepts was attached which showed semantic relevance of time concepts in the Holy Quran.

2.5.3 Legal domain ontology

Another retrieval system that relies on an ontology was introduced by Zaidi, S. and Laskri,(2008), to serve the legal domain, this system starts with the user keywords query, it can handle Arabic queries by translating them using wordnet, after that the query is expanded and formalized using a legal domain concepts ontology in order to enhance the next retrieving step accuracy, in the end this query is runned against a normal information retrieving system that retrieves answers from legal related documents.

2.5.4 Quran themes classification

An approach to build a quranic ontology was introduced in Ta'a et al, (2013) work, they build a theme representation for the Holy Quran not an actual data representation,

[7] http://corpus.quran.com/ontology.jsp

meaning the Holy Quran was modelled according to its structure and sections, not the actual concepts and their relations, furthermore no retrieval, question answering layer was introduced whatsoever nor a semantic layer, this work was just a simple representation for the Holy Quran structure in English language.

2.5.5 Quranic verse extraction

In this research Yauri et al, (2013) introduced a model of quranic ontology that is derived from previous built class hierarchy based ontologies, in this system no question answering capabilities is embedded it's a search engine that queries an ontology of quranic concepts.

2.5.6 Ontology Based Semantic Search in Holy Quran

Khan et al, (2013) introduced another very interesting research to model the animal related concepts in the Holy Quran into an ontology, this work was more than a class hierarchy like other Quranic ontology attempts, it clearly indicated relations other than the is-a or has-a relation, in this work researchers made a simulation on how queries derived from natural language questions can retrieve specific direct answers from a Quranic ontology, this work was very specific to animals in the Holy Quran and no other concepts was mentioned or used.

2.5.7 Comparison of Arabic and Quran ontologies

Table 3 - Comparison of Arabic and Quran ontologies

Criteria / Tool	Covered semantics	Relation diversity	Question Answering	Ontology coverage	Language used	#Entities/ #Relations	Format
Quran corpus	Yes	Poor	No	Mid	English	240/1	XML
Quran time	Yes	Poor	No	Poor	Arabic	51/7	OWL
Legal domain	Yes	Mid	No	Mid	Arabic-English	N/A	N/A
Quran themes	No	Mid	No	Mid	English	N/A	OWL
Quranic verse	No	Mid	No	Mid	English	650/300	OWL
Semantic search	Yes	Mid	No	Poor	English	N/A	OWL
AQO	Yes	Wide	Yes	Wide	Arabic	380/50	OWL

2.6 Ontology extractor from text

2.6.1 Ontology Extraction Approach for Prophetic Narration

A system that uses an automatic extraction method to acquire ontology from the Quran and Hadith domain text was developed by Fouzi et al, (2013), from the Ontology Research Team, UITM. This approach relied on a predefined model and concepts that will receive the concepts from any given document and place them as individuals into their already built ontology structure, we didn't use this model because it is not a general model that can be used on any Arabic text, it is limited to the Quran and Hadith domains, also no semantic or grammatical work was done to handle the Arabic words that can be syntactically similar but hold different semantics.

2.6.2 Text-To-Onto

Alexander and Raphael (2001), In this ontology extractor the researcher combines datamining techniques to detect concepts to build an ontology, they used association rules to detect concept concurrence within a document, this creates a class hierarchy of concepts not a real semantic ontology that goes beyond is-a or has-a relations.

2.7 Related work discussion and analysis

As for the related works that we mentioned we have to admit that in the previous years there was a very huge leap in the domain of ontology for the Arabic language in different aspect, one of the missing research aspects was that these ontologies in most of the cases had no semantic layer to enhance search result retrieval chances, we assume that by adding a semantic synonym generation layer to the search process over an ontology we will enhance retrieval.

Another key feature in search engines is question answering techniques, this aspect was totally absent in quranic ontology based systems, even though this domain (Holy

Quran) is highly targeted by questions not words, in our approach we implemented a question answering layer to fulfil this need in this domain.

in the domain of ontology automatic building approaches we noticed that almost all of the approaches uses data manning techniques such as association rules in order to detect keyword concurrence within a text, this is not efficient to build strongly semantic ontologies, because this doesn't detect all concepts in a particular domain, it just detects the concepts that are mentioned the most, whereas in some cases main concepts are mentioned once or twice within a domain document, another insufficiency s that these approaches doesn't detect semantic relations, it just detects concepts and their (is-a) and (has-a) relations but no other types of relations is extracted, let's not forget that the strength of the ontology data structure concept is being able to define all kinds of relations, to solve this issue we used a gummer driven approach in our relation extractor, the ontology building phase in our relation extractor depends on the semantics and grammatical structure of the targeted text, so it's more semantic and smart than datamining approached that relies only on word counts throughout a document.

After examining currently used Qur'an search engines, it was clear that all either didn't implement or poorly implemented these aspects:

- Semantically correlating concepts and features through the Qur'an verses.
- Absence of a strong word manipulation layer.
- Concentrating on syntax rather than semantics.
- Absence of a strong, full and meaningful ontology layer.
- Total absence of a question answering system.

2.8 Arabic language case

Current question answering or search approaches that is handling Arabic language are copying the same approaches that were used in English based systems, we believe that this was the main reason for the low accuracy these systems suffered from, Arabic has a unique structure that governs its syntax and semantics in a more sophisticated and formal manner due to the usage of diacritics, in the following tables (4+5) we will show how Arabic diacritics effects the meaning or even the grammatical form of words using the

same exact syntax, on the long run in this thesis we are trying to use approaches that can detect these semantic effects that these diacritics have on Arabic words.

Table 4 - Arabic diacritics effects on the meaning

Word	1st meaning	2nd meaning	3rd meaning	4th meaning
فرق	فَرِقَ: بكسر الراء: Panicked	فَرَقَ : بفتح الراء: Separated	الفَرَقُ : بضم القاف: the morning	الفِرْقُ : بكسر الفاء: portion slot.
الكفر	الكُفْرُ: بضم الكاف : Disbelief	الكَفْر: بفتح الكاف: To cover	والكَفِرُ : بكسر الفاء: great man	والكَفَرُ: بفتح الفاء: punishment
الخطبة	الخُطبة : بِضَمِّ الخاء : sermon	الخِطبة: بكسر الخاء : marriage proposal	الخُطْبَةُ بتشديد الطاء : Dusty color	-
العنان	العَنَان: بفتح العين : clouds	العِنان: بكسر العين : horse bridle	العِنان : بكسر العين: honorable	-
الفطر	الفَطْر: بفتح الفاء : crack	الفِطر: بكسر الفاء : breakfast	الفُطر: بضمِّ الفاء: mushrooms	الفِطر: بكسر الفاء: Islamic feast

Table 5 - Arabic diacritics effects on the grammar

Word	1st form	2nd form	3rd form	4th form
فرق	فَرِقَ : بكسر الراء verb	فَرَقَ : بفتح الراء verb	الفَرَقُ: بضم القاف noun	الفِرْقُ : بكسر الفاء noun
الكفر	الكُفْرُ : بضم الكاف verb	الكَفْر: بفتح الكاف verb	والكَفِرُ: بكسر الفاء noun	والكَفَرُ : بفتح الفاء ، noun
الفطر	الفَطْر : بفتح الفاء noun	الفِطر : بكسر الفاء verb	الفُطر: بضمِّ الفاء noun.	الفِطر : بكسر الفاء noun

2.9 Opportunities for more research

In this research we are implementing our approach on the Quranic text which is a special case or a subset from the Arabic language; we choose the Quranic text to implement this approach as a starter before tackling the general case (Arabic language documents) because this text (the Holy Quran) is a static text that didn't change since over than 1430 years, so it is a good case to test our approach on, because no additions or changes will happen during or after the development phase.

In order to take the semantic aspects to an advanced level, we will try to enhance our system with some of the Quranic exegesis (tafseer) that implemented the concept of semantics on the verse level, also we will try to add some machine learning aspects to our system.

It is worth mentioning that the understanding of the Holy Quran evolves with time, as more and more facts are explained by scientific discoveries and experiments. Our approach only attempt to enhance finding answers from the Holy Quran and it is only an experiment. Any incorrect results are unintentional and we apologize for them.

2.10 Summary for thesis related work

Research in this thesis will be mainly in providing the general public with a smart semantic Arabic question answering system, this will be done through enforcing the use of semantics and ontology concepts, along with other NLP techniques.

The research in this domain is very active but still slowly emerging regarding using the semantic power and ontology structure that is stored within the Holy Quran.

As for our data resources we will use the Quranic script as a main resource of data and information for our research, other resources will be the Arabic dictionaries where we will deduce synonyms in order to aid the semantic feature in our system. Finally, we will evaluate our model by comparing its results with already widely used online Quranic search engines.

2.11 Key concept definitions

Table 6 - Key concept definitions

Concept	Definition
Ontologies	**Data structures that are used to describe a domain by connecting concepts and instances with relations.**
Sparql query	**The type of query model needed to query data from ontologies**
Protégé	**A software used to build and edit ontologies.**
Concepts\classes	**Key indicators that identifies any given domain.**
Relations\axioms	**Connectors that semantically connect concepts.**
Individuals	**Instances of classes, they can benefit from relations as well.**

CHAPTER THREE: APPROACH

3.1 Introduction

In this research our main goal is to develop and implement an approach that can give us a semantic relation based structure to hold Arabic text into a semantic structure rather than a raw text format, this will give current search methodologies a significant leap, the current state of the data structure holding text targeted by search engines is limiting any improvement chances for any upcoming smarter search methodologies, current state of targeted data for search engines is raw streams and strings of data, this state can't indicate relations or any semantics between related concepts within this text, and this feature is a core feature for any approach based on semantics, question answering and information retrieval.

The new era of search engines as Google's knowledge graph by Singhal A., (2012), indicates that semantic based search engines that maps meaning to paragraphs and words, this will enhance search usefulness in the terms of better understanding for the users query semantics rather than its syntax only.

3.2 Why ontology?

In order to get this desired semantic structure we used the concept of ontology, the term ontology has its origin in philosophy and has been applied in many different ways. The word element onto- comes from the Greek ("being", "that which is"). The core meaning of ontology in the computer science field is a model for describing a particular domain that consists of types, properties, and relations (Garshol, (2004). There is also generally an expectation that the features of the model in an ontology should closely resemble the real world.

What many ontologies have in common in both computer science and in philosophy is the representation of entities, ideas, and events, along with their properties and relations, according to a system of categories.

Ontologies are considered by many the best approach to semantically model a domain. In ontologies we can create any desired axioms (constraints, relations) to connect related

concepts together, no expandability limitations and amazingly flexible structure where at any point you can extend or edit with no schema restrictions or normalization issues.

Let's say that we have 100 objects all are of the same class but each have different types of attributes, any database approach should consider all attributes for all instances either they need it or not, this is not efficient when dealing with poorly structured domains such as historical data for instance.

Extendibility is a major advantage that comes along if we considered an ontology structure to represent data, in an ontology structure you can simply connect new concepts, new relations, sub domains or even a whole related domain to your work, where as in other structures you have to edit and normalize the currently used schema for your data which is overwhelming and time consuming or even not even possible in some cases.

3.3 Building the ontology

In this research work we are also proposing a relation extractor model which can build up ontologies from Arabic texts by detecting concepts and relations between these concepts, Wong et al, (2012) was one of the main researchers to define this process using the term "ontology learning", they said: "ontology learning from text is the process of identifying terms, concepts, relations, and optionally, axioms from textual information then using them to construct and maintain an ontology for a specific domain".

Ontology building for our domain was handled using our relation extractor, some manual enhancement on the ontology was made later on, due to law accuracy, because of the fact that the proposed relation extractor still needs more work in order to handle all Arabic language grammatical cases and challenges previously shown in tables (4+5).

As a part of our research we built the first Arabic Quranic ontology (AQO). We have chosen Holy Quran as a test target for our system because of its unique structure and semantic behaviour of its verses and topic organization, Holy Quran is one of the unique texts that contains amazing amount of concepts and information within a relatively small amount of text.

In order to build the ontology we used Protégé ontology editor[8], we started detecting and fetching main concepts and relations from the Holy Quran then we connected them using protégé, our Arabic ontology for the Holy Quran contains around 380 concepts and 50 relations listed in (Appendix A,B,C), and to our knowledge, we are the first to propose and implement this contribution, this ontology is expected to be extended by more data in the near future, (Figure 1) shows how our ontology is used to aid the question answering process in our system.

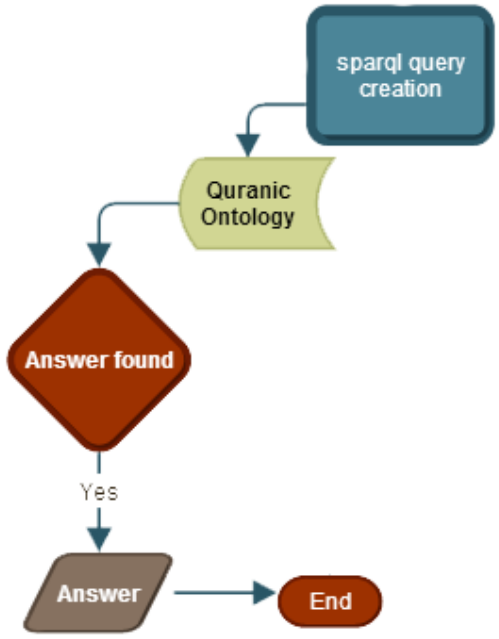

Figure 1 - How ontology aids question answering

The construction of the AQO greatly shifts the problem of question answer from simple text search to a richer semantic search that can utilize relationships between different text components. Now we have concepts that follow a class hierarchy and have semantic relation based axioms connecting them, having these features gave us the advantage of implementing more sophisticated approaches such as concept driven search and question answering system.

[8] http://protege.stanford.edu

3.4 Building semantic layer

One of the main issues when dealing with search methodologies is the ability to understand the semantics of a user's query, one of the major pillars of this process is being able to detect the meaning that the user is looking for instead of detecting it in a syntactic way, this means if the user is looking for a particular concept within a text but he only knows one word to describe this concept, the targeted text contains the intended concept but in another meaning, search engines or question answering systems with no semantic layer will not be able to retrieve any result whereas in a case of having a semantic layer it will be able to do that because it treats the keyword as a meaningful token not just a meaningless string of characters, meaning is given to words simply by generating its synonyms this will improve chances of finding a concept by its corresponding meaning rather than searching for the exact match for it, this is the desired result for most search engine users simply because in most cases they are not looking for the syntax instead they are looking for the semantic.

To implement the semantic feature for our system we used data from the website "MAANY"[9] the online Arabic language services website, we used their synonym dictionary to generate needed synonyms for keywords within our system.

Figure-2 shows how keywords synonyms are generated in the semantic layer of our system. The semantic layer gets keywords as an input, first it checks if these keywords are in our Holy Quran ontology if yes then just pass them to be searched in the next layer (ontology) if the keyword isn't in the ontology then generate its synonyms, after this synonym generation phase we will check, are any of these synonyms mentioned in the ontology, if yes then pass them to the next layer (ontology).

[9] http://www.almaany.com/

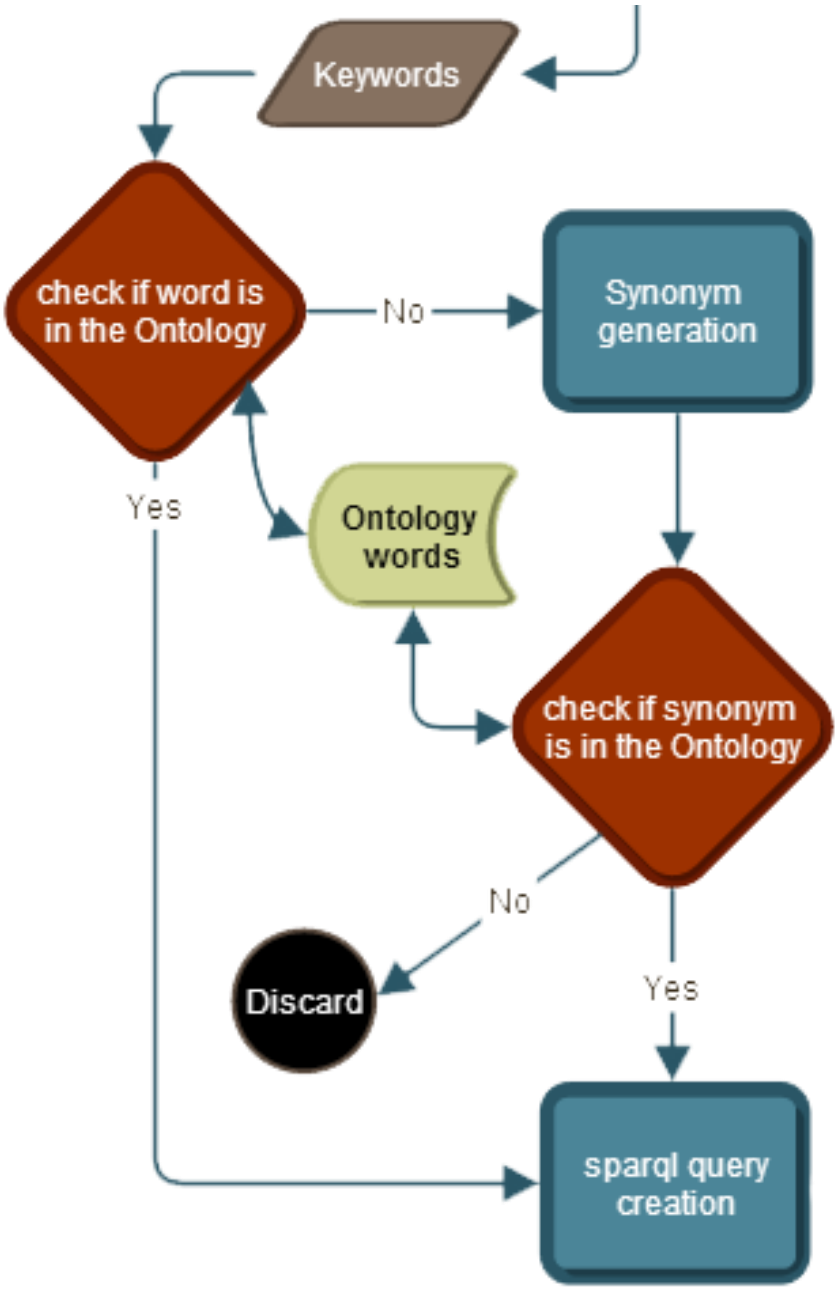

Figure 2 - synonym generation in the semantic layer

3.5 The question analyser layer

The third layer in our system is the question answering layer, in this layer we will analyse the question asked by the user, first we clean it by removing special characters(e.g. !,@,",& .. etc.) then we used the Stanford parser[10] to get grammatical tagging for the sentence, then we remove stop words listed in (Appendix F) and detect predefined domain identifiers that are listed in table 7, domain identifiers are used in order to know what kind of answer domain we are looking for, then we pass the remaining concepts with their grammatical tags to the next layer (semantic).

Table 7 - predefined domain identifiers

Domain identifier	Corresponding meaning in English
من	Who: this indicates a question asking about persons.
ما	What: this indicates a question asking about things.
أين	Where: this indicates a question asking about a place.
متى	When: this indicates a question asking about time.

Grammatical tagging is important because it will indicate the searched for query, as we said earlier in this research that we consider a verb to be the relation and the nouns to be the concepts, so in order to build up a meaningful efficient query we need to be aware of the grammatical statues of the user keywords, (Figure 3) shows the question analyser and its outputs.

[10] http://nlp.stanford.edu:8080/parser/index.jsp

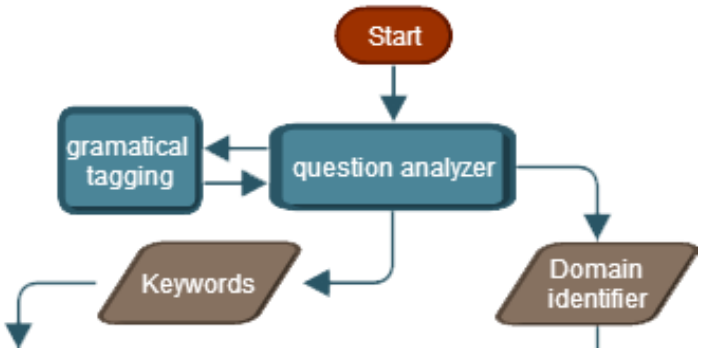

Figure 3 - The question analyser

3.6 Relation extractor

A very important evolving contribution that we introduce in this research is the relation extractor, to our knowledge there is no semantically based work done in this area for the Arabic language.

The need for this type of extractors emerged because of the rapidly evolving research in the field of ontologies, and as we mentioned before the main component and strength of an ontology based structure are relations that connect concepts together, so in order to describe any targeted text or domain; we have to extract relations and concepts from current raw unstructured text, to do that we proposed our relation extractor which relies on some assumptions.

In this extractor we assume that a relation relies on 2 major concepts, the relation (edge) and the concepts (nodes), also we assume that the relation in most of the cases is a verb, and the concepts are in most cases nouns, so by detecting these 2 main tags from the grammatically tagged question we can detect relations in the question that will be searched for in the ontology.

3.7 Current detected patterns

The main proposed patterns that are detected by our relation extractor are 3 patterns based on the location of the verb within a sentence.

1. The first pattern (Figure 4) is having 2 nouns followed by a verb, in this case the relation is (noun1) is related to (noun2) with the relation (verb).
2. The second pattern (Figure 5) is having a verb followed by 2 nouns.
3. The third pattern (Figure 6) is having a verb between 2 nouns.

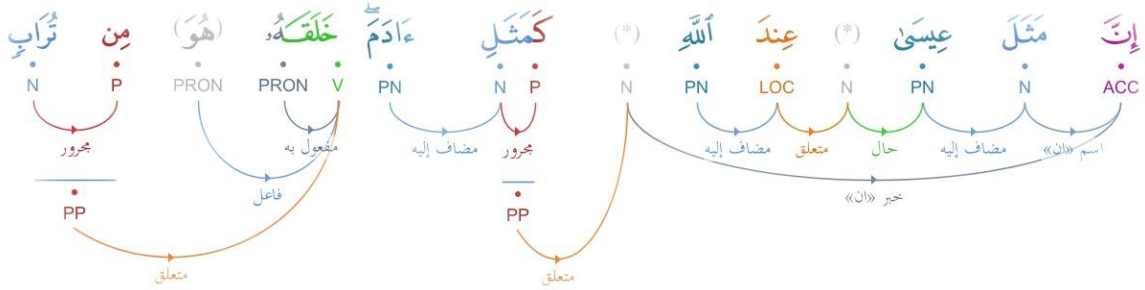

Figure 4 - Nouns are followed by a verb

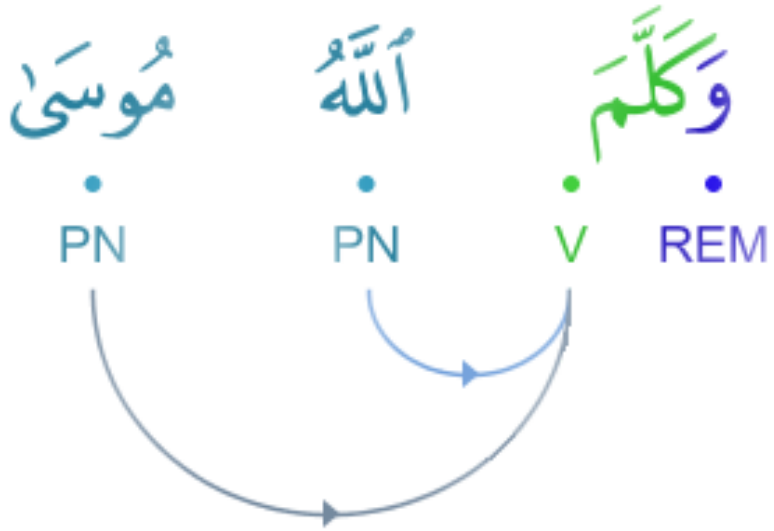

Figure 5 - Verb is followed by nouns

Figure 6 - Verb is between nouns

In (Figure 7) the reader can track inputs and outputs and observe how the relation extractor works its way starting from raw text to end up with an ontology.

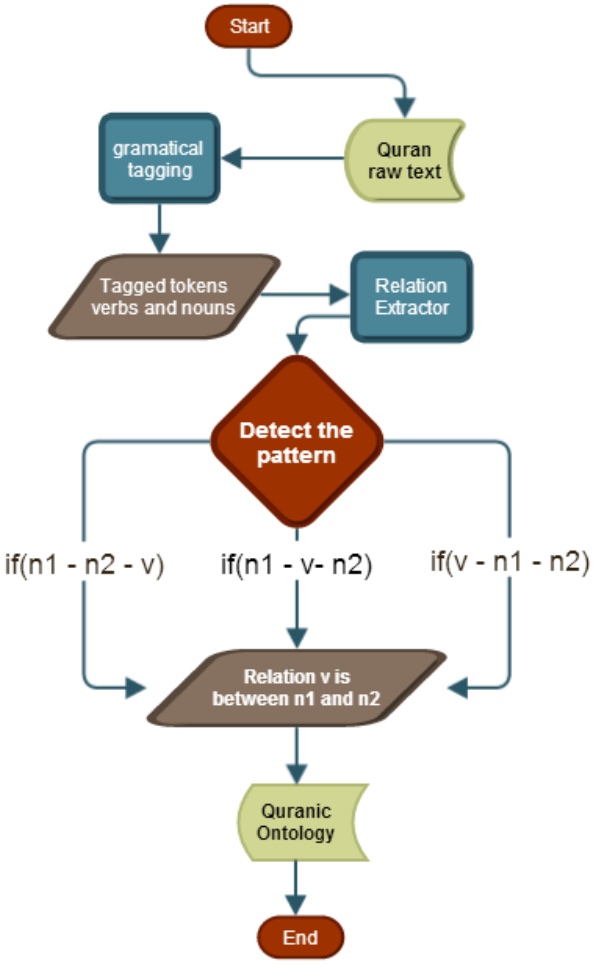

Figure 7 - The relation extractor flowchart

3.8 Current limitations

After observing the Arabic language general structure we have set some limitations for our relation extractor, due to the fact that the structure of Arabic language is highly complicated that it needs further analysis in order to reach the level of covering all of its cases with our extractor, after testing our relation extractor we detected some errors and low accuracy with detecting relations, mostly these errors and low accuracy was caused by noun based sentences and grammatically complex phrases, this will be solved in the future by enhancing the predefined relation patterns, due to the previously mentioned drawbacks, some manual editing was done on the ontology after running the relation extractor, in order to get better accuracy and less errors which will give us a better ontology to run our approach on, this law accuracy was a result of NLP perspectives and challenges in the Arabic language. Furthermore, it is noticed that our relation extractor is still limited to verb sentences. Currently we are using the Stanford parser to determine verbs and nouns in order to detect relations, more work should be done regarding the sentence norm structure in Arabic language where this will be a future work for this system.

3.9 Solution outline

As we mentioned earlier the architecture of our ontology driven question answering system starts from the process of building a domain ontology using a relation extractor witch will build the targeted ontology for our targeted text, then this ontology is used to implement our 3 layered system on, first we receive the question from the user, this question is analysed, main keywords are detected along with their grammatical tagging and the domain identifier, this directed to our semantic layer which will generate synonyms in order to enhance the chances of retrieving results from the ontology, then these keywords\synonyms are fetched into the sparql query builder which creates the suitable query to run against the ontology according to the grammatical tagging coming with the keywords, (Figure 8) shows an overall system flowchart.

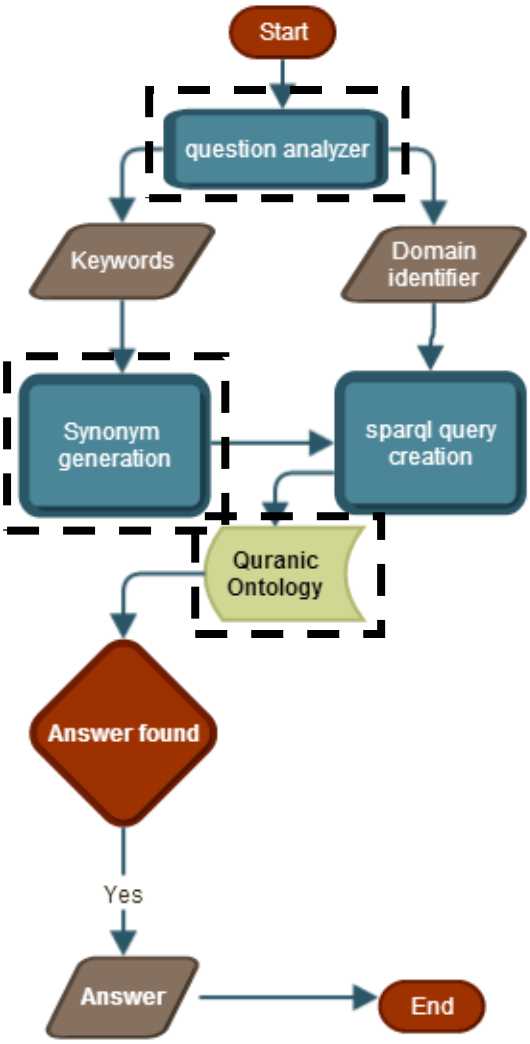

Figure 8 - The overall system flowchart

CHAPTER FOUR: IMPLEMENTATION DETAILS AND FEATURES

4.1 The ontology

During our research we worked on implementing our proposed system in order to demonstrate it, prove our assumptions and discover weaknesses, we started building the quranic ontology using protégé the ontology editor, we started by creating a class hierarchy to give some base structure for our ontology, then we collected concepts and their relations from the Holy Quran, fetching them into protégé and connecting them according to the quranic scripture, we reached a level of 380 concepts and 50 relations which we think is enough to test the system, as we previously said ontologies are easy to expand and add to, so building this ontology can go further from here, (Figure 9+10) shows some examples from our Quranic ontology, more examples can be found in (Appendix D).

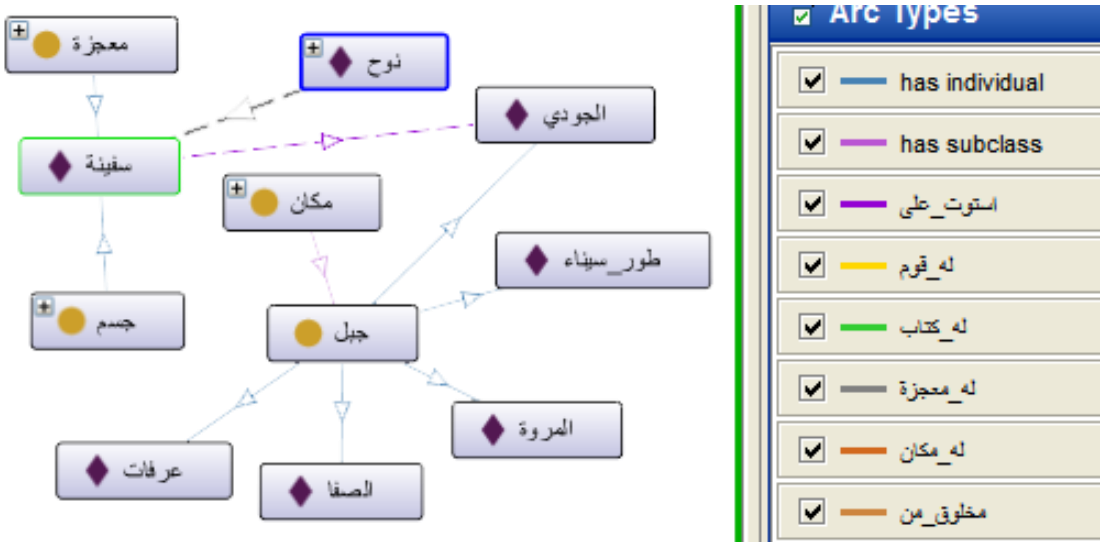

Figure 9 - Detecting concepts related to mountains in the Quran

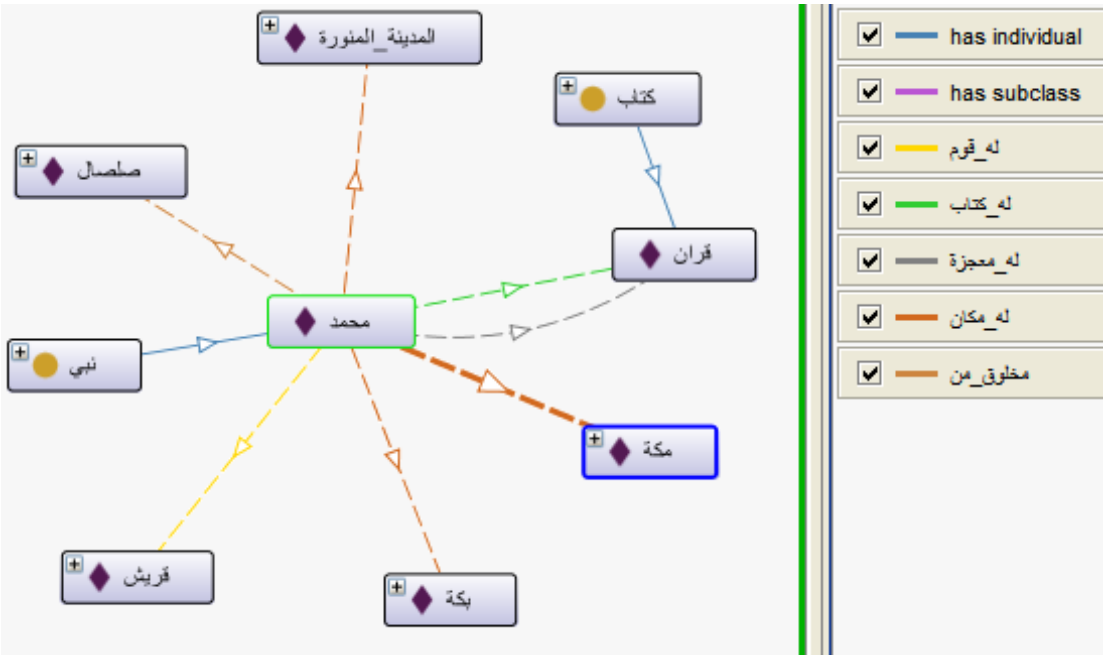

Figure 10 - Detecting concepts related to the prophet Mohammed PBUH

After building this ontology we were able to convert questions written in Arabic into sparql queries and run them against the ontology in order to retrieve answers.

In (Figure 11) we used sparql query to get names of mountains in the Holy Quran, another form of sparql query was used to get the miracle of the prophet Noah PBUH which returned “ship” as shown in (Figure 12).

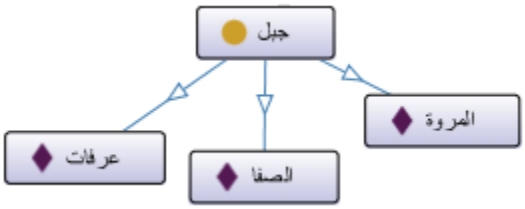

Figure 11 - Sparql query to get names of mountains in the Holy Quran

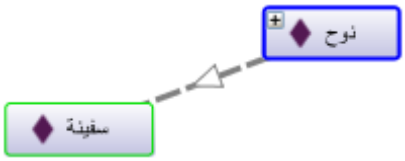

Figure 12 - Sparql query to get the miracle of the prophet Noah PBUH

Sparql queries are used to retrieve data from ontologies by using concepts and relations between them to drill-down to the desired result, some sparql query examples are shown in (Figures 13-15) they show how natural language question can be translated into sparql queries that can be used to retrieve information from an ontology.

```
من هو النبي الذي التقمه الحوت
"SELECT ?اسم WHERE {
? اسم rdf:type :نبي.
?اسم ? التقم: حيوان.
?حيوان rdf:type :حوت }"
```

Figure 13 - Sparql query example 1

```
ما هو عذاب قوم نوح؟
"SELECT ?عذاب
WHERE{?قوم :عذاب لهم "قوم نوح"
?عذاب rdf:type ?عذاب}"
```

Figure 14 - Sparql query example 2

```
ما هي الحيوانات المحرم أكلها؟
"SELECT ?حيوان
WHERE {?حيوان حرم أكل: thing:}
?حيوان rdf:type ?حيوان}"
```

Figure 15 - Sparql query example 3

4.2 The semantic layer

To implement the semantic layer we used datasets from “MAANY” the online Arabic language services website, we used their synonym dictionary data to create a database of Arabic words and their synonyms, and the used query to retrieve synonyms is:

Select * from [synonyms] where [word] like ('%var');

Some examples on the synonym generation process outputs are presented in (Table 8):

Table 8 - Synonems for some words

Arabic word	Synonyms	English word
رسول	مِرْسَال ، مُوفَد ، مَبْعُوث ، مُرْسَل، بَعِيث : مَبْعُوث : جَيش	prophet
جنة	مِجَنّ ، وِقَايَة ، دَرِينَة ، سِتْر ، دِرْع، حَدِيقَة : فِرْدَوْس ؛ دَار السَّلاَم	paradise
سفينة	فُلْك ، زَوْرَق ، قارِب ، باخِرَة ، مَرْكَب	ship
عين	مُقْلَة ، طرْف ، ناظِرَة ، باصِرَة ، جاسُوس ، مُرَاقِب ، عَمِيل	eye, spy
عقل	رَزَانَة ، وَقَار ، رَشَاد ، رَصَانَة ، فَهِمَ ، فَطِنَ ، أَدْرَكَ ، تَفَهَّمَ	wise, to realize

The input for this layer will be the keywords derived from the previous question analyzing layer, these keywords are the main words in the user question, they will be used to create the sparql query that will retrieve the answer from the ontology, in order to enhance our chance in retrieving an answer we use the semantic layer to give synonyms that indicate the same meaning of the keyword, in this case if the exact match for the word isn’t found then we can at least search for a semantically similar word to it, the output of this layer will be words that will build our intended sparql query.

The process starts with eliciting the keywords from the question analyzer, we will check if these words exist in our ontology, if yes then there is no need to process them with this layer, elsewise we will take the keyword and select its synonyms from our database, then

we will check if any of these synonyms exist in our ontology, if yes it will be passed to the next layer, elsewise it will be neglected.

4.3 The question analyser

In order to implement the question analyzer we built an ASP.NET webpage that will receive the users question and analyze it, to get grammatical tags for the user question to use them in next phases, we embedded the Stanford parser with our question analyzer, the Stanford parser is used to analyze sentences and give them grammatical tags.

The first step was to detect the domain of the question targeted answer ex. The word “when” indicates asking about time, and the word “where” indicates asking about a location and so on, this helps us when searching the ontology because it narrows our searched data to needed classes only.

After detecting the domain we start with the process of removing stop words, we created a file with common stop words (shown in Appendix F) in the Arabic language to ignore them from the keyword set when tokenizing the question.

The output of this layer is the keywords and the domain identifier, the keywords are passed then to the next layer, whereas the domain identifier is passed to the ontology layer directly.

CHAPTER FIVE: RESULTS AND EVALUATION

5.1 Introduction

Current systems suffer from the exact term driven search problem, where the search totally rely on the exact searched keyword syntactic form, and does not take care of the semantic side, in this system we solved this issue by using a synonym generating layer and an ontology layer, where we moved from searching strings of text into searching a semantic structure. Another aspect that this work solved, is the absence of an accurate question answering system that can answer factoid questions accurately, Currently there is no accurate semantic question answering system for the Holy Quran, due to this fact we are the only approach that solved the problem of question answering from the Holy Quran, nevertheless we can compare our system with current released online quranic search engines to show and evaluate the search ability for our first 2 layers, as for the question answering layer as we said we don't have any current systems solving the problem of Arabic question answering from an ontology nor from the Quran to compare with, inspite of that, we will use information retrieval accuracy measures to evaluate our system semantic search capabilities.

5.2 Evaluation

In (Table 9) we will show the results that current quranic search engines gave when we searched them for some keywords, N\A means no answer N\A+S means no answer but gives suggestions, N\R means not relevant.

Table 9 - Search capability evaluation for current Quranic search engines

Keyword / Search engine	التعليم education	مسيحي Christian	ميراث heritage	قارب boat	جهنم hell	معجزة عيسى miracle of Jesus	فيضان flood
quranicresearcher	N\A	N\A	2 - N\R	N\A	77	N\A	N\A
quranykeywords	N\A	N\A	2 - N\R	N\A	77	25 N\R	N\A
altafsir.com	N\A	N\A	N\A	N\A	77	N\A	N\A
quranix.net	N\A	N\A	N\A	N\A	77	25 N\R	N\A
quran.ksu.edu.sa	N\A	N\A	2 - N\R	N\A	77	N\A	N\A
alawfa.com	N\A	N\A	2 - N\R	N\A	77	25 N\R	N\A
quran.com	464	9	7	54 - N\R	77	N\A	N\A
alfanous.org	N\A	N\A+S	2 N\R	N\A+S	72	16 N\R	N\A+S
nss.cm	N\A	13	3	N\A	114	37	N\A
Our model	137	13	2	26	183	3	2

To detect and evaluate the accuracy of the retrieval aspect, a popular accuracy measure for retrieval systems is used, the precision and recall techniques, Precision percentage is the amount of relevant results united with the amount of retrieved results, this union over the amount of retrieved results, whereas Recall percentage is the amount of relevant results united with the amount of retrieved results, this union is over the amount of relevant documents. The main purpose of using the precision measure is to detect how many of the retrieved documents were actually relevant; on the other hand, the purpose of using the recall measure is to detect how many of the relevant documents were retrieved. Using a set of 35 randomly asked questions (Table 10), we found that the previously proposed approach retrieved the correct answer for 33 questions, leaving 2 questions unanswered, the unanswered questions either didn't have an answer in the Quran or was way complicated to be detected by our question analyser, precession for this test data was 94%, also recall was 94%. In the domain of retrieval of Holy Quran content this is considered very accurate and efficient when compared to current traditional Quranic keyword search engines.

$$Precision = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|}$$

$$Recall = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{relevant\ documents\}|}$$

Table 10 - Set of randomly asked questions to test retrieval accuracy

#	Arabic question	English translation	Correct answer	Retrieved answer
1	**من هي أم عيسى؟**	Who is the mother of Jesus?	مريم	مريم
2	**أين رست سفينة نوح؟**	Where did Noah's Ark dock?	الجودي	الجودي
3	**من هو محمد؟**	Who is Muhammad?	Full report	Full report
4	**ما هو عذاب قوم لوط؟**	What is the torment of sodomites?	خسف، صيحة	خسف، صيحة
5	**ما هو الحيوان الذي إلتقم يونس؟**	What animal swallowed Younis?	الحوت	الحوت
6	**ما هي الحيوانات التي حرم الله أكلها؟**	What are the animals that Allah forbid to eat?	Full report	Full report
7	**ما هو كتاب موسى؟**	What is the book of Moses?	التوراة	التوراة

8	من هو النبي الذي كتابه القرآن؟	Who is the prophet whom book is the Quran?	محمد	محمد
9	كم عدد آيات سورة آل_عمران؟	How many verses are there in Surat Al-Imran?	200	200
10	أين مكان نزول سورة البقرة؟	Where was Surat Al-Baqara revealed?	مدنية	مدنية
11	ماذا يحرم الإسلام؟	What Islam forbids?	Full report	Full report
12	من هو أخ مريم؟	Who is the brother of Mary?	هارون	هارون
13	ما علاقة آل فرعون ببني إسرائيل؟	What is the relationship between Pharaoh and the people of Israel?	Full report	Full report
14	من هم الملائكة في القرآن؟	Who are the angels in the Quran?	Full report	Full report
15	من هو أب مريم؟	Who is the Father of Mary?	عمران	عمران
16	ما هي الديانات في القرآن؟	What are the religions in the Quran?	Full report	Full report
17	ماذا خلق الله؟	Why did Allah create?	Full report	Full report
18	ما هي الحيوانات في القرآن؟	What are the animals in the Quran?	Full report	Full report
19	من ماذا خلق الجن؟	What material was the jinn created from?	نار	نار
20	ما هي معجزة نوح؟	What is the miracle of Noah?	السفينة، الطوفان	السفينة، الطوفان
21	ما هي مدينة محمد؟	What is the city of Muhammad?	مكة، المدينة	مكة، المدينة
22	ما هي نسبة الميراث للذكر؟	What percentage of the inheritance is for the male?	Variable, no fixed answer	N/A
23	ما هو عمر محمد؟	What is the age of Muhammad?	Not stated	N/A
24	ما هو كتاب عيسى؟	What is the book of Jesus?	الإنجيل	الإنجيل
25	من هو النبي الذي أمره الله أن يذبح ابنه؟	Who is the prophet that Allah had commanded him to slay his son?	إبراهيم	إبراهيم
26	من هم الأنبياء في القرآن؟	Who are the prophets in the Quran?	Full report	Full report
27	من هو النبي الذي كلمه الله؟	Who is the prophet that Allah talked to?	موسى	موسى
28	ما هي أركان الإيمان؟	What are the pillars of faith?	Full report	Full report
29	من هم الأقوام الظالمون في القرآن؟	Who are the bad tribes in the Quran?	Full report	Full report
30	من هم قوم النبي هود؟	Who are the people of Prophet Hood?	عاد	عاد

31	**من هو النبي الذي معجزته الناقة؟**	Who is the prophet whom miracle was the camel?	صالح	صالح
32	**من هم القوم الذين ذبحوا بقرة؟**	Who are the people who slaughtered a cow?	بني_إسرائيل	بني_إسرائيل
33	**من القوم الذين كان لهم كلب؟**	Who are the people who have had a dog?	أهل_الكهف	أهل_الكهف
34	**ما هي معجزات النبي عيسى؟**	What are the miracles of Prophet Issa?	Full report	Full report
35	**ما هو النبات الذي هزته مريم؟**	What is the plant that Mary rocked?	النخلة	النخلة

As for the relation extractor proposed model, to our knowledge there is no work done in this area, also our proposed model needs more work and enhancement regarding NLP perspectives because it is still limited to verb sentences, currently we are using the Stanford parser to determine verbs and nouns in order to detect relations, more work should be done regarding the sentence structure in Arabic language where this will be a future work for this system.. We will end this chapter by a conclusion describing the perspectives and future works.

5.3 Conclusion: perspectives and future works

In this research we have built an ontology based question answering system for the Arabic language, we choose the script of the Holy Quran to use as test data to test our system, also we proposed a relation extracting model to build Arabic ontologies that uses a different approach other than the currently used approaches, the combination of these techniques gave a semantic structure that gives meaning to strings, this opened the doors to implement a question answering layer that needs this semantic strength in order to function properly.

5.4 Summary of contributions (achievements and impact)

Our contributions in this research was building the first Arabic relation extractor, we needed to build this extractor to aid the evolving research area of ontologies and semantic

web, this extractor uses the grammatical structure of the Arabic language to detect relations and concepts that will be used to build an ontology.

Another contribution is building the first Arabic ontology for the Holy Quran, this ontology will open the doors wide for any future work in Quranic search approaches, and this structure shows and highlights the semantics that relies within the Holy Quran.

5.5 Outline open issues/directions for future work

For future work we are aiming to enhance our system in order to be able to adapt to any given Arabic based context; not only the Holy Quran, to do that we need to enhance the relation extractor by studying the Arabic language grammatical structure more in order to detect more accurate relations and cover more sentences other than verb sentences.

Furthermore we are looking for the new research area which is ontology based search, this area is a newborn evolving hot research topic, and in the future this will need more accurate Arabic ontology retrieval approaches based on quality ontologies build by accurate relation extractors.

Another evolving concern will be the Ontology migration/combination management procedure, ontology is the new trend in technology nowadays, Google and other main tech giants are starting to give big attention to this field, so the process of same domain ontology versions and distributions merging will be a hot research topic to work on.

REFERENCES

Abbas Noorhan, (2009). Qurany 'Search for a Concept' Tool and Website, Unpublished thesis, School of Computing, University of Leeds.

Alfaifi, AYG and Atwell, (2013). ES Arabic Learner Corpus v1: A New Resource for Arabic Language Research. In: Second Workshop on Arabic Corpus Linguistics, Lancaster, UK.

Al-Yahya Maha, Al-Khalifa Hend, Bahanshal Alia, Al-Odah Iman and Al-Helwah Nawal, (2010). An ontological model for representing semantic lexicons: an application on time nouns in the HOLY QURAN, the Arabian Journal for Science and Engineering, Volume 35.

Ash-Shanqeetee Muhammad Al-Ameen, (1955). Adwaa-ul-Bayaan fee Tafseer-il-Qur'an bil-Qur'an, Medina.

Atwell E., Brierley C., Dukes K., Sawalha M. and Sharaf A. M., (2011). An Artificial Intelligence Approach to Arabic and Islamic Content on the Internet, National Information Technology Symposium (NITS), Riyadh, Saudi Arabia.

Bdour Wafa N, Gharaibeh Natheer K., (2013). Development of yes/no Arabic question answering system, international journal of artificial intelligence & applications (IJAIA), Vol.4, No.1, January 2013.

Bekhti Smain, Rehman Amjad, Al-Harbi Maryam and Saba Tanzila, (2011). AQuASys: An Arabic Question-Answering system based on extensive question analysis and answer relevance scoring, International Journal of Academic Research, Vol. 3, No. 4.

Bernstein Abraham, Kaufmann Esther and Kaiser Christian, (2006). "Ginseng: A Guided Input Natural Language Search Engine for Querying Ontologies", Jena User Conference, Bristol, UK.

Brini Wissal, Ellouze Mariem, Mesfar Slim and Belguith Lamia, (2009). An Arabic Question-Answering system for factoid questions, IEEE International Conference on Natural Language Processing and Knowledge Engineering.

Dukes Kais, Atwell Eric, (2012). LAMP: A Multimodal Web Platform for Collaborative Linguistic Analysis, Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12).

Ezzeldin Ahmed Magdy, Shaheen Mohamed, (2012). A survey of Arabic question answering: challenges, tasks, approaches, tools, and future trends, the 13th international Arab conference on information technology ACIT.

Hammo Bassam, Abu-Salem Hani and Lytinen Steven, (2002). QARAB: A Question Answering System to Support the Arabic Language, SEMITIC '02 Proceedings of the ACL 02 workshop on Computational approaches to Semitic languages, Pages 1-11.

Harrag Fouzi, Alothaim Abdulwahab, Abanmy Abdulaziz, Alomaigan Faisal and Alsalehi Salah, (2013). Ontology Extraction Approach for Prophetic Narration (Hadith)

using Association Rules, International Journal on Islamic Applications in Computer Science and Technology, Vol. 1, Issue 2, September 2013, 48-58.

Ibn-katheer Ismail, (1370). Tafsir of Ibn Kathir, Damascus.

Kanaan Ghassan, Hammouri Awni, Al-Shalabi Riyad and Swalha Majdi, (2009). A New Question Answering System for the Arabic Language, American Journal of Applied Sciences 6(4):797-805.

Khan Hikmat Ullah, Saqlain Syed Muhammad, Shoaib Muhammad and Sher Muhammad, (2013). Ontology Based Semantic Search in Holy Quran, International Journal of Future Computer and Communication, Vol. 2, No. 6, December 2013.

Lopez V., Pasin M. and Motta E., (2005). AquaLog: An Ontology-portable Question Answering System for the Semantic Web, European Semantic Web Conference, Greece.

Maedche Alexander and Volz Raphael, (2013). The Ontology Extraction & Maintenance Framework (Text-To-Onto), IEEE International Conference on Data Mining.

Mazari Ahmed, Aliane Hassina and Alimazighi Zaia, (2013). Automatic construction of ontology from Arabic texts, Proceedings ICWIT 2012.

Singhal A., (2012).Introducing the Knowledge Graph: things, not strings, Official Google Blog.

Ta'a Azman, Zainal Abidin Syuhada, Abdullah Mohd, Ali Abdul Bashah, and Ahmad Muhammad, (2013). Al-Quran themes classification using ontology, Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013.

Tartir Samir, Arpinar I. Budak, (2007). Ontology Evaluation and Ranking using OntoQA, International Conference on Semantic Computing, ICSC 2007.

Tartir Samir, McKnight Bobby and Arpinar I. Budak, (2009). SemanticQA: Web-Based Ontology-Driven Question Answering, In the 24th Annual ACM Symposium on Applied Computing, Waikiki Beach, Honolulu, Hawaii, USA.

Trigui, Hadrich-Belguith and Rosso, (2010). DefArabicQA: Arabic Definition Question Answering System, 7th International Conference on Language Resources and Evaluation, LREC-2010.

UNESCO, (2012). On the World Arabic Language Day, UNESCO, 1995–2012.

Yauri Aliyu, Abdul Kadir Rabiah, Azman Azreen and Murad Masrah, (2013). Quranic Verse Extraction base on Concepts using OWL-DL Ontology, Research Journal of Applied Sciences, Engineering and Technology 6(23).

Zaidi S., Laskri M.T., Bechkoum K., (2008). A Cross-language Information Retrieval Based on an Arabic Ontology in the Legal Domain, the international conference on signal-image Technology & Internet-based systems (SITIS'05), Morocco, 2008.

Appendixes

Appendix A

Figure-16 shows a summary of concepts\classes used in this Quranic ontology:

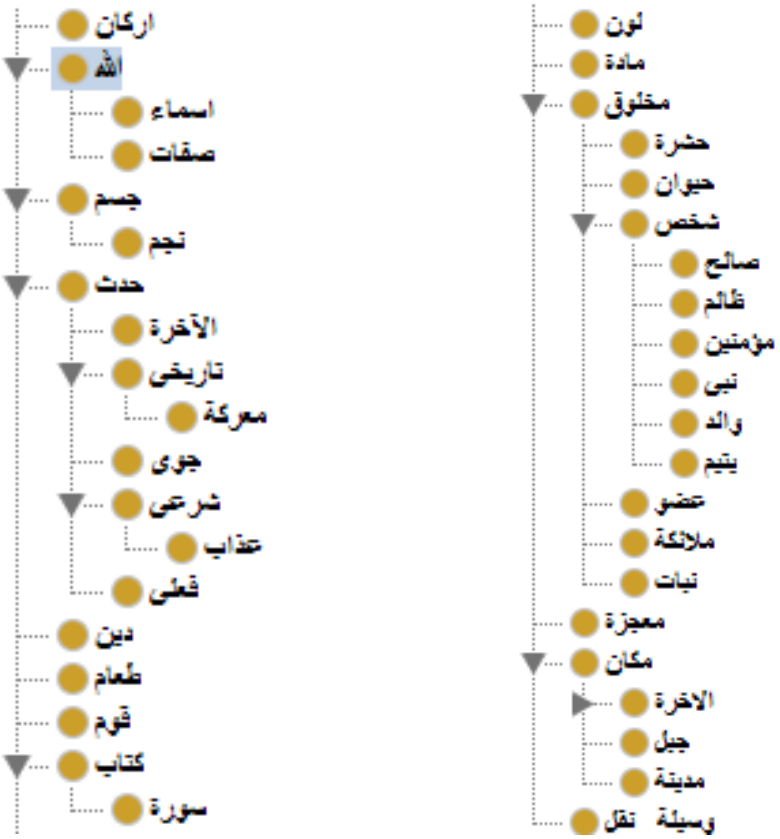

Figure 16- Concepts of the Quran ontology

Appendix B

Table-11 shows a summary of relations used in this Quranic ontology:

Table 11 - Relations used in the Quranic ontology

ركن	حرم	اسم	مكان	مكان_النزول	يذبحون_أبناء
صفات	خالق	ترتيبها	جائهم	له_سورة	إن_أنعمنا_عليه
طعام	خُلِقَ	حجم	نجّى	له_عدو	إن_مسه_الضر
ظلموا	ذبح	صفة	هز	له_قوم	عدد_آياتها
غرق	له	عدد	هو	له_كتاب	احل_لهم
اخت	أم	اتخذ	ملاحظات	له_معجزة	فرق_لهم
في	بنت	عمر	لهم_عذاب	له_مكان	استوت_على
قال	يعض	لون	مخلوق_من	نجى_من	يستحيون_نساء
السنة	الأفكار_العامة	سبب_نزول	نزلت_قبل	نزلت_بعد	مرتبطة_بالمعنى_مع

Appendix C

Table-12 shows a summary of entities used in this Quranic ontology:

Table 12- Concepts used in the Quranic ontology

آل_عمران	اخضر	تبع	زفير	قبلة	مجوسية	خمر	أصحاب_الحجر
آل_فرعون	ازرق	ترائب	زقوم	قثاء	محمد	خنزير	أصحاب_الرس
آل_لوط	اسد	تراب	زنجبيل	قران	مخاض	دم	أصحاب_الفيل
أبواب	اسلام	تغيظ	زيت	قرد	مدين	ذئب	أصحاب_الكهف
أجنة	اسماعيل	تين	زيتون	قريش	مرجان	ذباب	أصحاب_الأخدود
أذن	اسود	ثرى	سبأ	قلب	مرسلين	عين	أصحاب_الأيكة
أصابع	اصفر	ثمود	سجين	قمر	مريم	ذهب	الابل_من_الصخر
نمل	الأحقاف	ثوم	سحاب	قوم_نوح	مسيحية	نطيحة	الارض_المقدسة
نوح	نور	جالوت	سدرة	كعب	مصر	نمل	الايمان_بالرسل
نبي	هابيل	جبريل	سفينة	كلب	مطر	ملح	الايمان_بالقدر
نحلة	هاروت	جبهة	سقر	كوكب	نخلة	ملك	الايمان_بالكتب
نخلة	هارون	جرادة	سلسبيل	لؤلؤ	مكة	رحمة	الايمان_بالله
نطيحة	هامان	جمل	شجرة	لبن	عاد	النساء	الايمان_بالملائكة
أعراب	هلوع	يأجوج	شعيب	لحم	عدس	امعاء	الايمان_باليوم_الاخر
أعرض	هود	جناح	شفاء	يثرب	عذاب	انجيل	جنات_عدن
أنصار	ورقة	جنة	شفة	يد	عرفات	ايمان	إني_عبد_الله
إرم	البقرة	جهنم	شمس	لسان	عزير	آزر	يا_ليتني_اتخذت_مع_الرسول_سبيلا
إلياس	الجودي	حبة	صابئين	لظى	عسل	بابل	يا_ليتني_لم_اتخذ_فلانا_خليلا
إنسان	الروم	حجابا	صالح	لقمان	عظم	بحر	ياقوت
يئس	السامرى	حديد	صحراء	لوط	علق	بدر	يبرأ_الأكمه_والأبرص
ابراهيم	الصابئين	حرير	صلصال	مأجوج	عمران	برق	يد_الله_مغلولة
ابيض	الصفا	حصان	ضفدع	ماروت	عنب	بصل	يكلم_في_المهد
احمر	القمل	حنين	طالوت	ماعز	عنكبوت	بعوضة	طور_سيناء
ريح	المتقين	حوت	يونس	متردية	عيسى	بغل	بنى_إسرائيل
زبور	يثرب	حية	طوفان	مجرمين	ميكائيل	نبي	بني_ادم
الأحزاب	البلد	عبس	الشمس	الفتح	الليل	ص	النازعات
الأحقاف	البينة	الحديد	الشورى	الفجر	طه	الناس	المؤمنون
الأعراف	التحريم	الحشر	الصافات	الفرقان	المائدة	النبأ	الحجرات
الأعلى	التغابن	الدخان	الصف	الفلق	الماعون	النجم	الذاريات
الأنبياء	التكاثر	غافر	الضحى	الفيل	المجادلة	النحل	الشعراء
الأنعام	التكوير	الرحمن	الطارق	القارعة	المدثر	النساء	فصلت
الأنفال	التوبة	الرعد	الطلاق	القدر	ق	النصر	المرسلات
الإخلاص	التين	الروم	الطور	القصص	المزمل	النمل	قريش
الإسراء	الجاثية	هود	العاديات	القلم	المسد	النور	لقمان
الإنسان	الجمعة	الزلزلة	العصر	القمر	المطففين	الهمزة	المنافقون
الإنشقاق	الجن	الزمر	العلق	القيامة	المعارج	الواقعة	لحم_الخنزير
الإنفطار	الحاقة	السجدة	نوح	الكافرون	الملك	يس	العنكبوت
البروج	الحج	الشرح	الغاشية	الكهف	الممتحنة	إبراهيم	الزخرف
يوسف	الحجر	فاطر	الفاتحة	الكوثر	محمد	سبأ	آل عمران
زجاج	المروة	خبز	طير	مريم	ناصية	نحلة	مكانا_شرقيا

بكة	الميتة	خرطوم	ظالم	لحم_طير	يهودية	فردوس	ذبح_لغير_الله
ناصية	منخنقة	رعد	موسى	ذو_القرنين	يديه	فرعون	ربنا_آتنا_من_لدنك_رحمة
بقل	مهل	رمان	بقرة	ذو_الكفل	ميكائيل	فضة	ربنا_هيء_لنا_من_أمرنا_رشدا
مجوس	مالك	غنم					

Appendix D

The following figures shows the concepts and relations mentioned in (Table11 & 12) connected and presented using the (OntoGraf) plugin for Protégé:

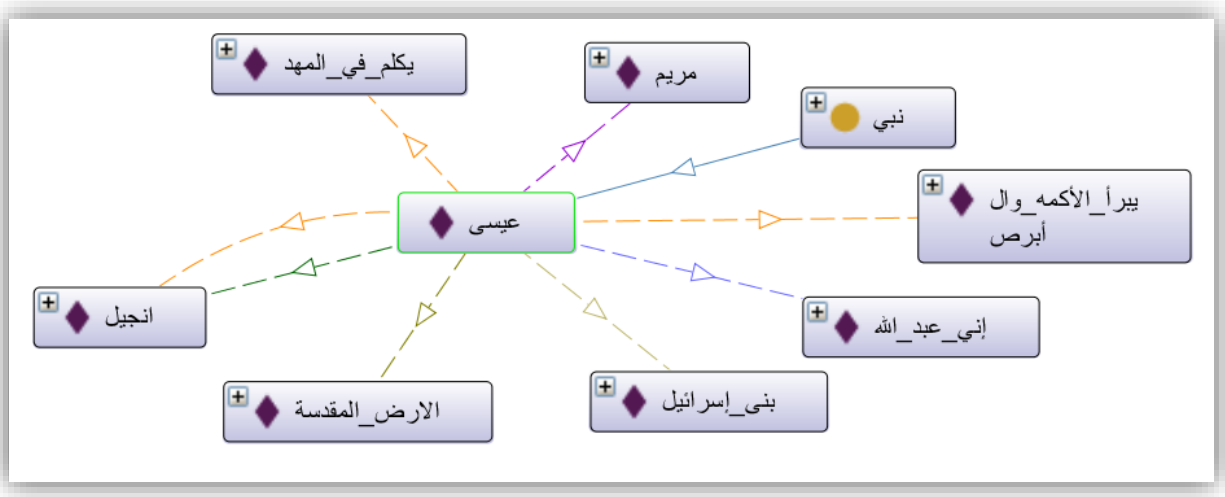

Figure 17 - The concept (عيسى), its relations and related concepts

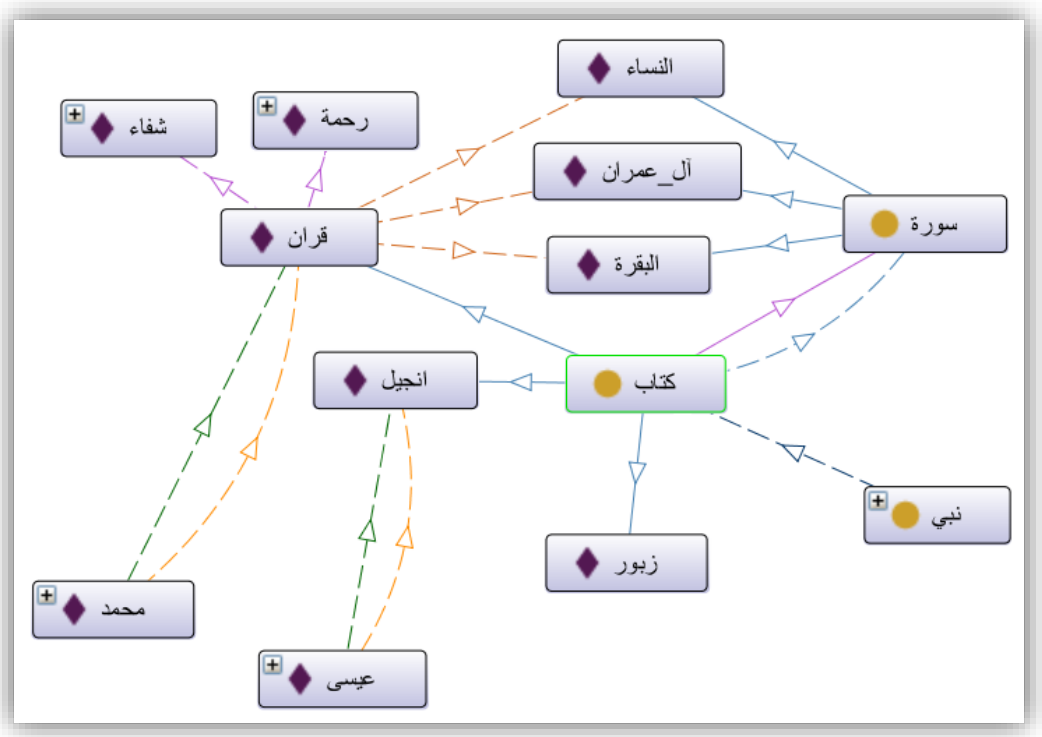

Figure 18 - The concept (كتاب), its relations and related concepts

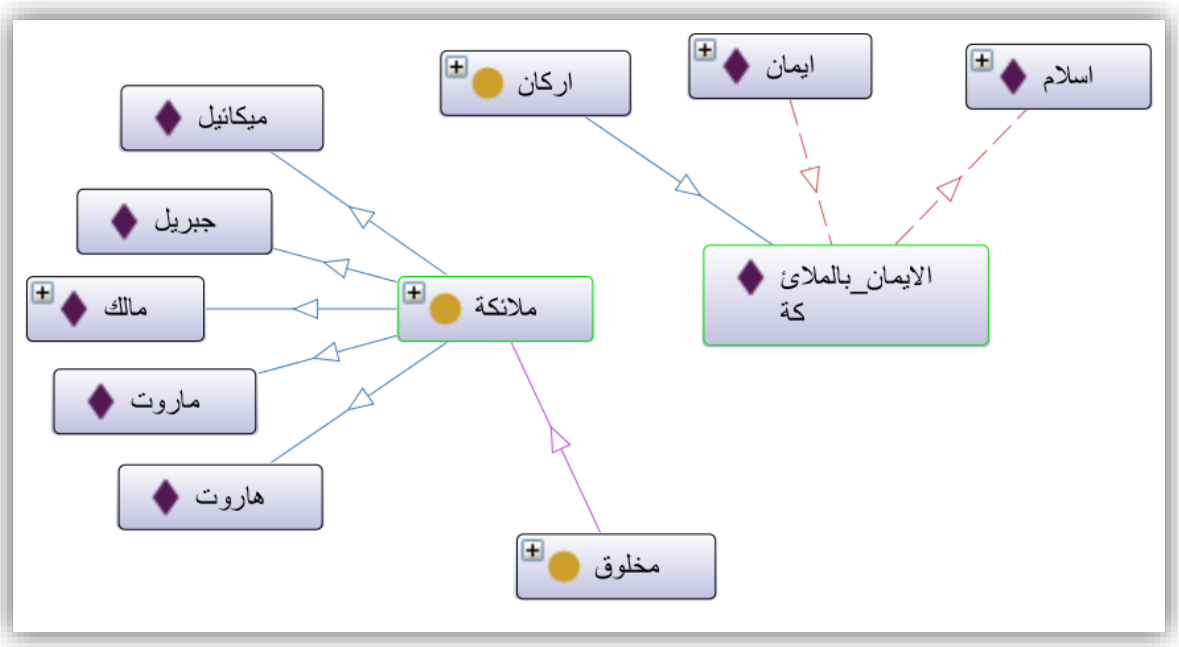

Figure 19 - The concept (ملائكة) , its relations and related concepts

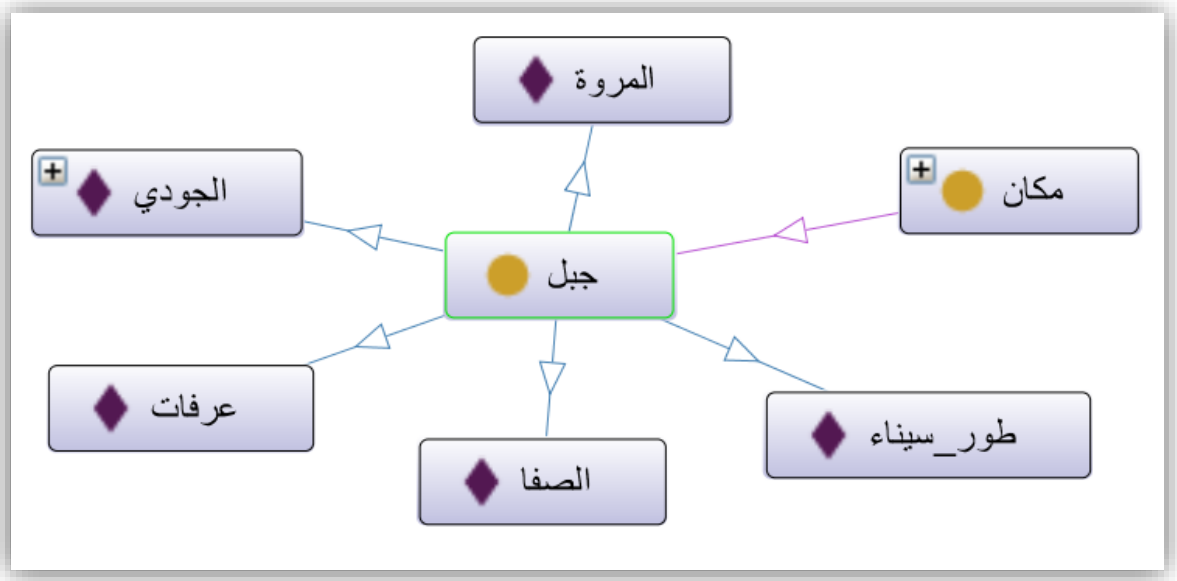

Figure 20 - The concept (جبل) , its relations and related concepts

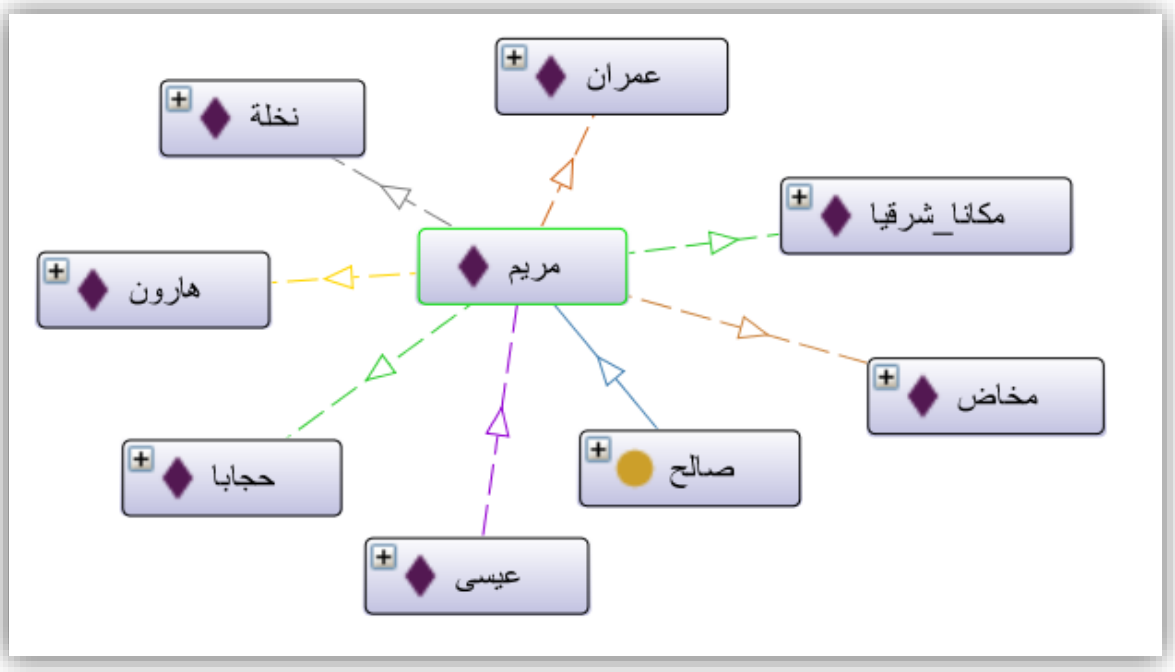

Figure 21 - The concept (مريم) , its relations and related concepts

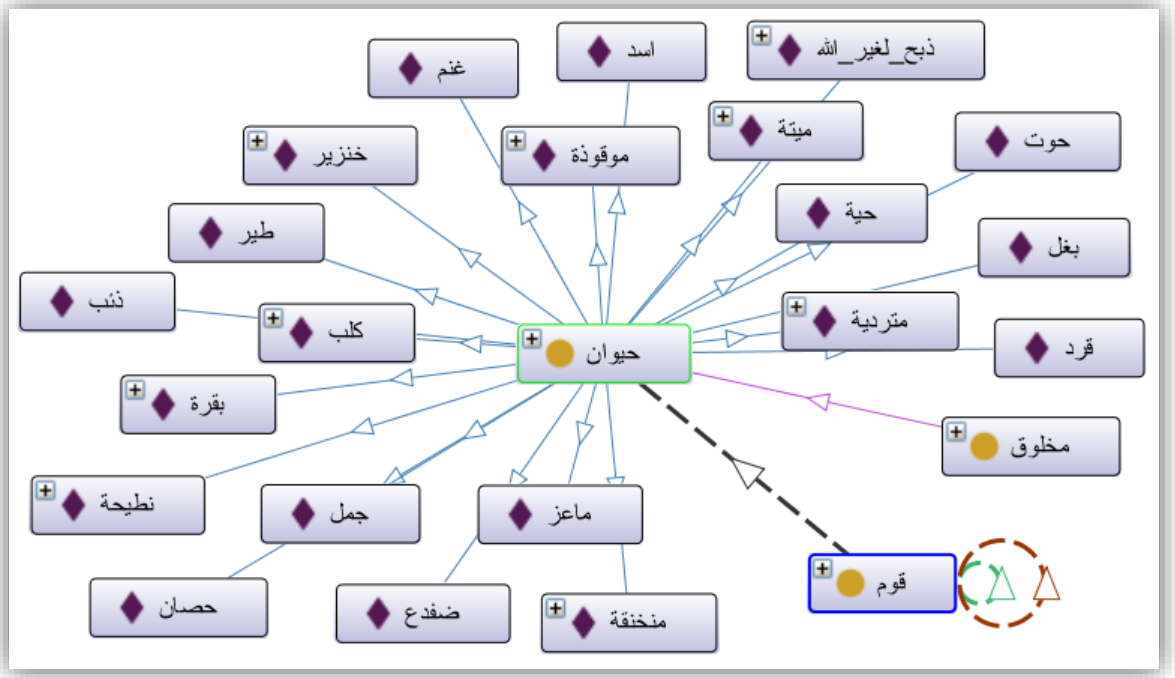

Figure 22 - The concept (حيوان) , its relations and related concepts

Appendix E

Here we have some code snippets showing the internal structure of the Quranic ontology OWL file:

- In this block the is-a relation (نبي\prophet) is used to define (هود) as a prophet.

```
<owl:NamedIndividual rdf:about="&ontologies;QURAN.owl#هود">
  <rdf:type rdf:resource="&ontologies;نبي"/>
</owl:NamedIndividual>
```

- In this block we can see how similar concepts using different terms can be described, this is done using the keyword (sameAs).

```
<owl:NamedIndividual rdf:about="&ontologies;QURAN.owl#جنات_عدن">
  <rdf:type rdf:resource="&ontologies;QURAN.owl#الاخرة"/>
  <owl:sameAs rdf:resource="&ontologies;QURAN.owl#جنة"/>
  <owl:sameAs rdf:resource="&ontologies;QURAN.owl#فردوس"/>
</owl:NamedIndividual>
```

- In this block we can see how the concept (بنى_إسرائيل) is fully described using the (is-a) relation and other relations such as (غرق، ذبح، نجى).

```
<owl:NamedIndividual rdf:about="&ontologies;QURAN.owl#بنى_إسرائيل">
  <rdf:type rdf:resource="&ontologies;قوم"/>
  <rdf:type rdf:resource="&ontologies;QURAN.owl#ظالم"/>
  <غرق rdf:resource="&ontologies;QURAN.owl#بحر"/>
  <ذبح rdf:resource="&ontologies;QURAN.owl#بقرة"/>
  <نجّى rdf:resource="&ontologies;QURAN.owl#عذاب"/>
</owl:NamedIndividual>
```

- In this block we can see how the concepts forbidden by Islam law, is fully described using the (حرم\forbid) relation.

```
<owl:NamedIndividual rdf:about="&ontologies;QURAN.owl#اسلام">
  <rdf:type rdf:resource="&ontologies;QURAN.owl#دين"/>
  <حرم rdf:resource="&ontologies;QURAN.owl#خمر"/>
  <حرم rdf:resource="&ontologies;QURAN.owl#خنزير"/>
  <حرم rdf:resource="&ontologies;QURAN.owl#دم"/>
  <حرم rdf:resource="&ontologies;QURAN.owl#ذبح_لغير_الله"/>
  <حرم rdf:resource="&ontologies;QURAN.owl#متردية"/>
  <حرم rdf:resource="&ontologies;QURAN.owl#منخنقة"/>
  <حرم rdf:resource="&ontologies;QURAN.owl#موقوذة"/>
  <حرم rdf:resource="&ontologies;QURAN.owl#ميتة"/>
  <حرم rdf:resource="&ontologies;QURAN.owl#نطيحة"/>
</owl:NamedIndividual>
```

- In this block we can see how the concept (آل_فرعون) is fully described using the (is-a) relation (قوم) and other relations such as (ظلموا ، يذبحون_أبناء ، يستحيون_نساء).

```
<owl:NamedIndividual rdf:about="&ontologies;QURAN.owl#آل_فرعون">
  <rdf:type rdf:resource="&ontologies;قوم"/>
  <ظلموا rdf:resource="&ontologies;QURAN.owl#آل_فرعون"/>
  <يذبحون_أبناء rdf:resource="&ontologies;QURAN.owl#بنى_إسرائيل"/>
  <يستحيون_نساء rdf:resource="&ontologies;QURAN.owl#بنى_إسرائيل"/>
</owl:NamedIndividual>
```

- In the following block we can see how a comment can be attached to a concept, in this case the concept is (**أجنة**).

```
<owl:NamedIndividual rdf:about="&ontologies;QURAN.owl#أجنة">
  <rdf:type rdf:resource="&ontologies;QURAN.owl#مخلوق"/>
```

```
الَّذِينَ يَجْتَنِبُونَ كَبَائِرَ الْإِثْمِ وَالْفَوَاحِشَ إِلَّا اللَّمَمَ إِنَّ رَبَّكَ وَاسِعُ الْمَغْفِرَةِ هُوَ أَعْلَمُ بِكُمْ <rdfs:comment>
إِذْ أَنْشَأَكُمْ مِنَ الْأَرْضِ وَإِذْ أَنْتُمْ أَجِنَّةٌ فِي بُطُونِ أُمَّهَاتِكُمْ فَلَا تُزَكُّوا أَنْفُسَكُمْ هُوَ أَعْلَمُ بِمَنِ اتَّقَى </rdfs:comment>
</owl:NamedIndividual>
```

- In this block we can see how the concept (مريم) is fully described using the (is-a) relation (صالح) and other relations such as (اتخذ، بنت، جاء، اتخذ، هز، اخت).

```
<owl:NamedIndividual rdf:about="&ontologies;QURAN.owl#مريم">
  <rdf:type rdf:resource="&ontologies;QURAN.owl#صالح"/>
  <اتخذ rdf:resource="&ontologies;QURAN.owl#حجابا"/>
  <بنت rdf:resource="&ontologies;QURAN.owl#عمران"/>
  <جاء rdf:resource="&ontologies;QURAN.owl#مخاض"/>
  <اتخذ rdf:resource="&ontologies;QURAN.owl#مكانا_شرقيا"/>
  <هز rdf:resource="&ontologies;QURAN.owl#نخلة"/>
  <اخت rdf:resource="&ontologies;QURAN.owl#هارون"/>
</owl:NamedIndividual>
```

Appendix F

Table-13 shows a summary of stop words used in the question analyzer:

Table 13- Stop words used in the question anylizer

وليس	به	يكون	وهو	حتى	من	في	الى	يلي	ضد	بعد
ليس	مساء	عن	لكن	وعلى	إن	عليها	فيها	وبين	التي	كذلك
لكنه	مع	دون	حول	عنه	ما	أي	وكانت	ليسب	لا	ومن
كان	نحو	لن	جدا	بين	قد	تكون	أنه	هذه	ثم	فقط
وثي	لدى	بد	كل	اللذين	عند	لو	ذلك	فيه	فإن	هؤلاء
علي	إذ	أو	لها	تحت	فهو	وفي	بها	منه	عنها	هو
إذا	إلى	منها	يوم	معه	قبل	هناك	أمام	لذلك	كانت	وقد
اصبح	مايزال	لايزال	لازال	مازال	إلي	الي	لا	ما	و	او
ليت	كأن	إن	ليس	صار	بات	ماانفك	مافتئ	مابرح	ظل	اضحى
وان	فانه	الذين	انه	اليها	بدلا	اي	ذات	وله	اول	ضمن
وأن	لدي	بهذا	يمكن	اليه	الذي	بإن	أبو	مما	ستكون	فكان
امسى	أمسى	أصبح	تلك	وكان	على	أحد	أين	اين	الذى	هن
ولايزال	لاسيما	لعل	بل	فقد	ومع	أن	حين	أما	الذي	منذ
لهذا	وهذا	والذي	هنا	كيف	كما	عليه	والتي	هذا	له	ولكن
آل	وأبو	وهي	اذا	هي	حيث	هل	لم	اليوم	لأن	لهم
الحالي	الا	الذي	أضحى							

ملخص

اللغة العربية هي اللغة الخامسة على المستوى العالمي لعدد المتحدثين بها، وهي تعتبر اللغة الثانية لأكثر من مليار مسلم حول العالم. بالرغم من هذا، مقدار البحث العلمي المبذول لتطوير منهجيات معالجة النصوص العربية عموما والقرآنية منها خصوصا، ما زال قليلا ولا يتناسب مع هذا الإنتشار العالمي لهذه اللغة.

أحد المنهجيات الرئيسية التي تتأثر باللغة في مجال البحث العلمي في علم الحاسوب هي قوة محركات البحث وإسترجاع المعلومات، فيما يخص اللغة العربية ما زالت محركات البحث في مرحلة البحث الحرفي عن النصوص، حيث أن النتائج المسترجعة من عمليات البحث تكون معتمدة إعتمادا تاما على نص الكلمات الحرفي التي قام المستخدم بإدخالها، فتقوم بإسترجاع النص الحرفي وتهمل الجانب المعنوي تماما.

أمر آخر مهم وهو طريقة تنظيم النصوص المستهدفة في عملية البحث، فما زالت هذه النصوص عبارة عن أحرف وكلمات خالية من الترابط المعنوي، حيث لا يمكن ربط المصطلحات بمصطلحات أخرى تشبهها معنويا أو موضوعيا، مما أدى إلى ضعف في منهجيات الإجابة عن الأسئلة وأنظمة الإسترجاع المعتمد على المعنى.

بالإعتماد على الحقيقة التي تقول أن اللغات بشكل عام يعتمد فهمها على المعنى لا على التركيبة الحرفية للكلمات، بمعنى أنه يمكن لكلمة واحدة أن تعطي أكثر من معنى حسب موضع ذكرها، أو حسب تشكيلها، هذا يدل على أن الكلمات والنصوص يجب أن يتم معالجتها بطريقة جديدة تراعي الجانب المعنوي للنصوص عوضاً عن التركيز على البنية الحرفية للكلمة أو النص.

بالإعتماد على ما سبق، نحن نقترح في هذه الرسالة منهجية لتخطي هذه المشاكل المتعلقة بالبنية التنظيمية للنصوص والمتعلقة بالدعم الدلالي والمعنوي لها، مما يمكننا من دعم وتحسين نوعية أنظمة ومنهجيات البحث والإجابة عن الأسئلة من النصوص العربية.

لحل مشكلة البحث الحرفي وتحويله لبحث معنوي قمنا بإستعمال مرادفات اللغة العربية لتوليد بدائل في البحث، أما بخصوص تحسين البنية التنظيمية للنصوص المراد البحث فيها، قمنا بإستعمال منهجية الإنتولوجيا لتحويل النص خالي المعنى إلى نص منظم ذو عناصر وعلاقات واضحة، بدلاً عن صورته اللامعنوية.

حتى نقوم بإختبار هذه المنهجية المقترحة قمنا بإختيار نص يعتمد كليا على المعنوية في تركيبه، ألا وهو نص القرآن الكريم. حاليا من يريد البحث في القرآن الكريم عن مواضيع محددة يجب عليه أن يستعمل في بحثه كلمات دقيقة ومطابقة تماما للكلمات المذكورة في القرآن حتى يستطيع إسترجاع نتائج، وهذا ليس هو البحث المطلوب غالباً خصوصاً في حالة البحث داخل نص يعتمد كليا على المعاني والإستدلالات والترابطات بين المواضيع والكلمات. البحث في القرآن الكريم يكون عادةً بحث عن مفهوم أو موضوع، لا عن نص حرفي. بكلمات أخرى، الباحث عادةً يبحث عن التشابه المعنوي لا عن التشابه الحرفي، ولهذا قمنا بإضافة نظام توليد المرادفات في منهجيتنا.

نقطة مهمة أخرى، وهي أن معظم عمليات البحث في القرآن الكريم يكون الهدف منها السؤال عن أمر معين لا البحث عن كلمة بحرفيتها، مثال(كم نام أهل الكهف؟)، حتى نستطيع إحتواء هذا النوع من المدخلات قمنا بإضافة نظام تحليل وإجابة على الأسئلة في منهجيتنا حتى نستطيع فهم سؤال المستخدم والإجابة عليه.

من الإضافات المهمة التي قمنا بتقديمها في هذه الرسالة هو، مستخرج العلاقات، والذي يقوم بإستخراج العلاقات من النصوص العربية إعتمادا على التركيبة النحوية للكلمات، وذلك لغايات إستعمال هذه العلاقات في بناء أونتولوجيا لهذا النص، مما يجعل النص منظم بطريقة معنوية حيث تكون كل المفاهيم فيه مرتبطة بالمفاهيم الأخرى عن طريق علاقات واضحة يمكن من خلالها تسهيل عمل منهجيات الإجابة عن الأسئلة والبحث المعنوي للنصوص العربية، وهذه المستخرج حسب علمنا هو أول مستخرج علاقات معنوي للغة العربية.

خلاصة الكلام، هذا البحث يقوم على بناء نظام إجابة أسئلة، يعتمد على نظام دعم معنوي يقوم بتوليد المرادفات، يعتمد على طبقة تنظيم معنوي للنص المراد البحث فيه بإستخدام الأونتولوجيا التي تم بنائها وتشكيلها عن طريق مستخرج علاقات معنوي، هذا النظام حسب علمنا سيقدم لأول مرة للغة العربية عموما، وللقرآن الكريم خصوصاً.

إستخدام المنهجيات الدلالية لخدمة أنظمة الإجابة عن الأسئلة باللغة العربية من القرآن الكريم

بواسطة
هاشم عدنان جدعان شميساني

بإشراف
د. سمير الترتير

قدمت هذه الرسالة استكمالاً لمتطلبات الحصول على درجة الماجستير في علم الحاسوب

عمادة البحث العلمي والدراسات العليا
جامعة فيلادلفيا

1435هـ