Zayed University

ZU Scholars

All Works

7-1-2018

Extracting semantic relations from the Quranic Arabic based on Arabic conjunctive patterns

Rahima Bentrcia
Université Hadj Lakhdar de Batna

Samir Zidat
Université Hadj Lakhdar de Batna

Farhi Marir
Zayed University

Follow this and additional works at: https://zuscholars.zu.ac.ae/works

Part of the Computer Sciences Commons, and the Social and Behavioral Sciences Commons

Recommended Citation

Bentrcia, Rahima; Zidat, Samir; and Marir, Farhi, "Extracting semantic relations from the Quranic Arabic based on Arabic conjunctive patterns" (2018). *All Works*. 1632.
https://zuscholars.zu.ac.ae/works/1632

This Article is brought to you for free and open access by ZU Scholars. It has been accepted for inclusion in All Works by an authorized administrator of ZU Scholars. For more information, please contact Yrjo.Lappalainen@zu.ac.ae, nikesh.narayanan@zu.ac.ae.

Journal of King Saud University – Computer and Information Sciences 30 (2018) 382–390

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Extracting semantic relations from the Quranic Arabic based on Arabic conjunctive patterns

Rahima Bentrcia [a,*], Samir Zidat [a], Farhi Marir [b]

[a] Department of Computer Science, Lastic Laboratory of the Systems and the information and Communication Technologies, Chahid Mostefa Ben Boulaid, University of Batna 2, Batna, Algeria
[b] College of Technological Innovation, Zayed University, Dubai, United Arab Emirates

ARTICLE INFO

Article history:
Received 29 June 2017
Revised 13 August 2017
Accepted 11 September 2017
Available online 12 September 2017

Keywords:
Ontology
Text mining
Arabic AND conjunctive patterns
Arabic grammar
Semantic relations
Quranic Arabic Corpus

ABSTRACT

There is an immense need for information systems that rely on Arabic Quranic ontologies to provide a precise and comprehensive knowledge to the world. Since semantic relations are a vital component in any ontology and many applications in Natural Language Processing strongly depend on them, this motivates the development of our approach to extract semantic relations from the Quranic Arabic Corpus, written in Arabic script, and enrich the automatic construction of Quran ontology. We focus on semantic relations resulting from proposed conjunctive patterns which include two terms with the conjunctive AND enclosed in between. The strength of each relation is measured based on the correlation coefficient. Finally, we evaluate the significance of this method by using hypotheses testing and Student *t*-test. The obtained results are very promising since we combine an accurate Arabic grammar with strong statistical techniques to prove the existence and measure the strength of this type of semantic relations.
© 2017 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

An early definition of the term ontology appeared in 1993 (Gruber, 1993), where it is defined as a specification of a conceptualization. Maedche and Staab stated the original description of ontology learning from a text as the acquisition of a domain model from data (Maedche and Staab, 2001), where the extracted knowledge from the text is represented by concepts and relationships. Hence, semantic relations are an important element in the construction of ontologies (Alvarez et al., 2007). Besides holding together the concepts that represent the domain, they solve the ontology structuring problems. Furthermore, providing richer semantics to these relations facilitates selecting the operations that can be performed on them and the task that the ontology can tackle. However, semantic relations have not been given the attention they deserve because of the difficulty to capture the whole information related to the problem domain as well as the different possible representations provided for a specific relation.

The process of ontology learning passes through several tasks organized in a layer cake. Each layer is explained deeply in (Cimiano, 2006; Liu et al., 2011). Traditionally, ontology construction depends on domain experts, but it is lengthy, costly, and controversial (Navigli et al., 2003). Therefore, automatic ontology construction approach was suggested but it is also still a difficult task due to the lack of a structured knowledge base or domain thesaurus (Lee et al., 2007).

On the other side, there is a lack in the developed approaches that deal with ontology learning from texts written in Arabic script due, for example, to the nature of Arabic writing, the semantic ambiguity of words, and the shortage in resources and tools that support Arabic (Farghaly and Shaalan, 2009). For Quran ontologies, all studies aim to achieve the purpose of understanding Quran as a source of knowledge and facilitating information retrieval automatically. Therefore, Quran can be presented to the world and employed very efficiently in many linguistic and religious studies. Currently, there are no complete Quran ontologies; many of them have covered specific topics in Quran or special types of words rather than the whole Quran words (Saad et al., 2010). Also, many researchers have built ontologies for parts of Quran and very few have used the entire Quran. Moreover, each ontology has focussed only on one or two types of relations between terms such as

* Corresponding author.
E-mail addresses: rahmabentrcia@yahoo.com (R. Bentrcia), samir.zidat@gmail.com (S. Zidat), Farhi.marir@zu.ac.ae (F. Marir).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2017.09.004
1319-1578/© 2017 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

synonymy and meronymy (Shoaib et al., 2009). As a validation method, the existing ontologies are evaluated by domain experts who rely on scholarly sources in their decisions (Ta'a et al., 2013).

Despite all these challenges, ontologies can provide potential benefits for a lot of applications such as text classification and clustering (Bloehdorn and Hotho, 2004), where additional conceptual features extracted from ontologies are used to enhance the bag-of-words model. In information retrieval and extraction, ontologies can solve the problem of vocabulary mismatch between documents and user queries, and many other problems (Guarino et al., 1999; Elabd et al., 2015). Also, ontologies provide the necessary concepts that represent a specific domain and state which semantic relations potentially hold between them (Nirenburg and Raskin, 2004).

For information systems that rely on Quranic Arabic, Quranic ontologies can be exploited very efficiently in improving the performance of the information extraction process. One such system is the question answering system for Quran (Q&A system) (Abdelnasser et al., 2014; Hakkoum and Raghay, 2016) where the user enters a question in Arabic related to Quran and the system retrieves the most relevant answer from Quran. Conventional Q&A systems are based on simple keyword search to find the best answer for the user query without knowing the meaning of the words and expressions used and the relationship between them. This kind of search has many limitations especially when applied to the Quranic Arabic where for example polysemy and synonymy are very common. Therefore, using ontologies to present the Quran knowledge provides semantics, logic, inference, and deep reasoning to the Q&A systems so complex and ambiguous queries could be solved very accurately.

In this paper, we introduce a novel approach that aims at enriching the automatic construction of Quran ontology. We extract the whole relations that exist in the conjunctive phrases which can be defined as any two words combined by the conjunction AND. The two words could be nouns, proper nouns, or adjectives. The main contribution is that we define a hybrid method to extract semantic relations from Quran based on strong and solid rules. First, we exploit an efficient rule in Arabic grammar, which is AND conjunction, to extract several types of semantic relations. AND conjunction is a well-known grammatical tool that combines terms which have a degree of association between each other. The proposed set of patterns is used to extract the AND conjunctive phrases from the Quranic Arabic Corpus and not to extract a specific type of semantic relations, as is known in the pattern-based methods. Second, we use an accurate measurement, namely the correlation coefficient, to find the association value between the combined words. This is totally new and useful in this field and different from other common measurements such as Mutual Information (MI) and t-score. Finally, we combine statistical tests (testing hypothesis and student *t*-test) and domain experts to validate the results achieved. All the reported approaches in the field of Quran mining (Alrehaili and Atwell, 2014) depend on either domain experts or exegesis books such as Tafsir of Ibn Kathir in the validation process. This step is very essential and cannot be neglected because the holy Quran is a very sensitive and critical text. Basically, we present a scientific validation approach to consolidate the domain experts' decisions.

Manually, we reveal three different categories of semantic relations from the entire Quran based on one type of Arabic patterns which is the conjunctive patterns. The rationale behind selecting conjunctives in general and the conjunction AND in particular stems from its importance in Arabic grammar. It is considered as the basic Arabic conjunctive for its frequent use and probable indications for different meanings in the sentence (AL-Taweel, 2009). We initiate the proposed work by converting the Quranic Arabic Corpus from Buckwalter transliteration scheme to Arabic script. Then, we extract the set of words that form ontology terms. We cover nouns, proper nouns, and adjectives. Next, we apply a set of conjunctive patterns, which we previously defined, to extract the candidate relations from the corpus. A filtering approach based on a correlation coefficient is used to select strong relationships. Finally, we validate the proposed approach by using testing hypothesis and Student *t*-test besides domain experts. This statistical technique supports the output of domain experts in order to give the most accurate and correct results.

The rest of the paper is organized as follows. Section 2 presents related work. Preprocessing phase is described in Section 3 and ontology learning from the holy Quran phase is introduced in Section 4. Section 5 presents the validation phase. Experimental results are discussed in Section 6 and conclusion is presented in Section 7.

2. Related work

Ontology learning from text in general occupies a large area in computer science, whereas ontology learning from the holy Quran in particular suffers from the specific nature of the Arabic script and the depth of knowledge needed in this field (Habash, 2010). However, few recent Quranic studies were interested in developing approaches which accomplish ontology learning tasks and represent the Quranic knowledge in a semantic way as sets of concepts and relations. Duckes initiated the Quranic Arabic Corpus (QAC) which is the first online collaboratively constructed linguistic resource with multiple layers of annotation including part-of-speech tagging, morphological segmentation and syntactic analysis using dependency grammar (Dukes and Habash, 2010; Dukes and Atwell, 2012). Also, the author built ontology from (QAC) which finds relations between proper nouns or any nouns if they represent well-defined concepts such as the names of animals, locations, and religious entities. The ontology was validated based on scholarly sources, namely Tafsir of Ibn Kathir (Ibn Kathir, 1999).

A large corpus named QurAna (Sharaf and Atwell, 2012) was created from the original Quranic text, where specific types of words are considered as ontological concepts. Personal pronouns are extracted and tagged with their antecedents. These antecedents are maintained as an ontological list of concepts which improves information systems performance. Abbas exploited an existing index of Quranic topics from a scholarly source: Tafsir of Ibn Kathir, to develop Qurani (Abbas, 2009), which is a tool looks for concepts in the holy Quran and provides English translations for the verses containing these concepts. Yauri et al. proposed a system that reused Leeds ontology (Quranic Arabic Corpus ontology) to model Quran domain knowledge, using Web Ontology Language OWL (Yauri et al., 2012). The system added the act concepts related to specific topics in Quran such as praying, Zakat, sin and rewards, and showed the relations between them using Description Logic. The user of this model can semantically retrieve important concepts from the holy Quran. Verses referring to particular concepts could also be retrieved. Ul Ain and Basharat introduced DataQuest (Ul Ain and Basharat, 2011), an efficient framework for modelling and retrieving knowledge from distributed knowledge sources primarily related to the holy Quran and scholarly texts, with the use of semantic web, information extraction, and natural language processing techniques. The documents are annotated using the domain ontology. Thus, users can query that filtered and concise knowledge using a semantically based intelligent search engine. Another work which covers a specific topic in the holy Quran was conducted by Al-Yahya and colleagues to build a computational model for representing Arabic lexicons using ontologies (Al-Yahya et al., 2011). The model has been implemented on the Arabic language vocabulary associated with "Time nouns" vocabulary in the holy Quran. The ontology consists of 59 words; only 28 of them are used as a basis for the model design and organized semantically into a hierarchical classification with general concepts at the top, and specific concepts at the bottom. In addition, Baqai et al. developed knowledge based platforms that used semantic web technologies to model, store, publish, reason and retrieve knowledge from distributed sources related to the holy Quran and associated scholarly texts (Baqai et al., 2009).

In the proposed work, the framework for extracting semantic relations from the holy Quran using Arabic 'AND' conjunctive was built in three distinct phases: pre-processing, ontology learning from Quran, which includes terms and relations extraction, and validation, as clarified in Fig. 1. The following sections discuss these phases in detail.

3. Pre-processing phase

Finding an available, robust, and accurate part-of-speech tagging system to process Quranic Arabic texts is not an easy task since it deals with a sacred text and a language with a complex morphological structure. For the sake to describe the Quranic corpus for all readers, the Quranic Arabic Corpus (Dukes and Habash, 2010) is an integrated and reliable linguistic resource which consists of 77,430 words of Quranic Arabic, divided into 114 documents. Each word is tagged with its part-of-speech as well as multiple morphological features that are based on the traditional Arabic grammar. Also, it is stored as a text file and is available for free.

The data in the corpus is written in Buckwalter Arabic transliteration scheme and organized into four columns as follows:

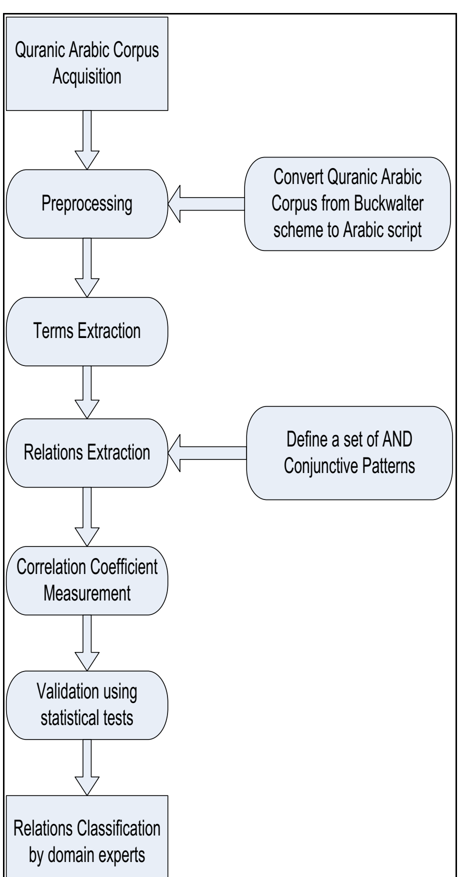

Fig. 1. Ontology Learning from the Holy Quran Phases.

1. LOCATION: consists of four parts: (chapter no: verse no: word no: part no).
2. FORM: consists of the main parts of the word.
3. TAG: includes the part-of-speech tag for each part of the word such as Noun, Verb, Adjective, etc.
4. FEATURE: describes morphological features of the word such as Root, Stem, Gender, etc.

Fig. 2 shows a verse of Alikhlas chapter, in Buckwalter transliteration scheme.

Because Buckwalter transliteration scheme is not an easy way for users to read and understand the corpus, we need a simple but an efficient pre-processing method to represent it in a clearer and more readable format such as the Arabic script. Therefore, we develop a conversion method to transfer back each character from Buckwalter scheme to its equivalent Arabic character. These include Arabic alphabet and diacritics. We prefer to use a straightforward algorithm to accomplish this task instead of using a sophisticated tool, so that acceptable performance is reached with less computational costs.

Fig. 3 demonstrates a sample of the Quranic Arabic Corpus converted to Arabic script.

4. Ontology learning from the holy Quran phase

Ontology learning from a specific text means extracting the main terms and relations that represent the domain (Liu et al., 2011). This phase includes two major components: terms extraction and relations extraction. Also, we define another component that seeks extracting conjunctive patterns to accomplish the relation extraction task, as detailed in Section 4.2.

4.1. Term extraction

We start this phase by extracting terms from the Quranic Arabic Corpus, including nouns, proper nouns, and adjectives in their stem form, to avoid considering different forms of the same word as different multiple words. For the sake of clarity, we use the word 'term' to refer to the words that exist in the ontology, and the word 'word' elsewhere. Quranic Arabic Corpus is a text file organized into four columns and many rows, as depicted in Section 3. To access the file contents, we use the ordinary read/write/search file functions to read the file line by line, search the TAG column looking for words with POS tagging equal to noun, proper noun, or adjective. The stem forms of the resulting words are extracted from the FEATURE column and stored as strings of characters. Finally, we remove the duplicated stems and store the unique ones in the term list. Arabic diacritics are characters like letters and we use the same functions to manipulate them. They are very important elements that construct Arabic words and distinguish a word's meaning from that of other words with similar patterns of letters. For these reasons, we do not remove the diacritics. Rather, we exploit the POS tagging information available in the corpus and

```
LOCATION          FORM     TAG      FEATURE
(112:1:1:1)       qulo     V
          STEM|POS:V|IMPV|LEM:qaAla|ROOT:qwl|2MS
(112:1:2:1)       huwa     PRON
          STEM|POS:PRON|3MS
(112:1:3:1)       {ll~ahu  PN
          STEM|POS:PN|LEM:{ll~ah|ROOT:Alh|NOM
(112:1:4:1)       >aHadN   N
          STEM|POS:N|LEM:>aHad|ROOT:AHd|M|INDEF|N
OM
```

Fig. 2. Sample of the Quranic Arabic Corpus in Buckwalter transliteration.

```
(112:1:1:1)          قُلْ          V
          STEM|POS:V|IMPV|LEM:قَالَ|ROOT:qwl|2MS
(112:1:2:1)          هُوَ          PRON
          STEM|POS:PRON|3MS
(112:1:3:1)          اللَّهُ          PN
          STEM|POS:PN|LEMاللَّه|ROOT:Alh|NOM
(112:1:4:1)          أَحَدٌ          N
          STEM|POS:N|LEM:>أَحَد|ROOT:AHd|M|INDEF|NO
M
```

Fig. 3. Sample of the Quranic Arabic Corpus converted from Buckwalter transliteration to Arabic.

Table 1
The Arabic conjunctions mentioned in the holy Quran.

Conjunctive	ثم	أو	و	بل	الفاء	أم
	Then	Or	And	But	Then	Or

we deal with the stem form of words. As an example, consider the two Arabic Quranic words ('الجَنٰة' and 'الجِنٰة'), which have identical letters, different diacritics, and hence different meanings. Based on their stem form ('جَنٰة' and 'جِنٰة') respectively, we have two different words and not one, which would not be the case if we removed the diacritics. For Quranic text, taking the diacritics into consideration with an accurate manipulation is very important and increases the efficiency of the proposed approach.

One basic text mining technique to process textual data is to convert each word in the text into a numerical value that represents word importance in the corpus (Weiss et al., 2005). We achieve this goal by constructing a matrix called term-document matrix where its rows are the extracted Quranic terms and its columns are the Quranic chapters (i.e. surahs). Each term has a specific weight in each document in the corpus. There is an efficient statistical method to calculate weights called Term Frequency Inverse Document Frequency (tf.idf) (Salton and McGill, 1986):

$$\boldsymbol{wij} = \boldsymbol{tf}_{ij} . \boldsymbol{log}\left(\frac{\boldsymbol{N}}{\boldsymbol{df} i}\right) \tag{1}$$

where $w_{i,j}$ is the weight of term i in document j, $tf_{i,j}$ is the number of occurrences of term i in document j, N is the total number of documents in the corpus, and df_i is the number of documents containing term i. A high weight in tf.idf is reached by a term with a high frequency in a given document and a low document frequency in the corpus; the weights hence tend to filter out common terms which are less discriminative.

The result of this process is a matrix of 3267 rows of unique words, namely terms, and 114 columns of Quranic chapters (i.e. surahs). The elements of this matrix are the calculated tf.idf weights of each term in a given chapter.

4.2. Conjunctive patterns extraction

Arabic grammar is very rich of patterns and clauses that serve different purposes in the sentence. In this work, we call a conjunctive pattern every two terms enclosing AND conjunction in between. The considered terms could be nouns, adjectives, or proper nouns. There are nine conjunctions in Arabic, where the two combined terms must have a type of association between each other. However, only six of them have a conjunctive role in the holy Quran, and are repeated several times (Adhima, 1972), as shown in Table 1.

Before elucidating the conjunctive-based relations, we define a set of conjunctive patterns/ rules based on a deep study of Arabic grammar (Al-Zujaji, 1984; Al-Ghalayini, 2007), POS tagging, and morphology features found in the Quranic Arabic Corpus. We treat only the cases where the two combined terms may be nouns, proper nouns, or adjectives. Other complex cases are beyond the scope of this work because they need specific knowledge resources such as exegesis of the holy Quran. This set is as follows:

1. Noun + Conjunctive "*AND*" + Noun: this pattern is for extracting any two nouns with AND conjunctive in between, such as: 'هُدًى و مَوْعِظَةً', which means guidance and instruction. Different cases of this pattern are explained bellow:
   a. Noun + Conjunctive "*AND*" + Noun + Determiner "ال" + Noun: the two combined nouns are followed by a third noun which starts with a determiner, like 'إِحْسَانًا و ذِي الْقُرْبَى', which means good and to relatives.
   b. Noun + Conjunctive "*AND*" + Noun + Noun: the two combined nouns are followed by a third noun, like 'فِتْنَةً و ابْتِغَاءَ تَأْوِيلِهِ', which means discord and seeking an interpretation.
   c. Noun + Conjunctive "*AND*" + Noun + Determiner "ال" + Adjective: the two combined nouns are followed by an adjective which starts with a determiner, like 'الأَرْضُ وَ السَمَاوَاتُ الْعُلَى', which means earth and highest heavens.
   d. Noun + Conjunctive "*AND*" + Noun + Adjective: the two combined nouns are followed by an adjective like 'بُهْتَانًا و إِثْمًا مُبِينًا', which means injustice and manifest sin.
2. Adjective + Conjunctive "*AND*" + Adjective: this rule is for extracting any two adjectives with AND conjunctive in between, such as 'ثَيِبَاتٍ و أَبْكَاراً', which means previously married and virgins.
3. Proper Noun + Conjunctive "*AND*" + Determiner "ال" + Proper Noun: this rule is for extracting any two Proper nouns with AND conjunctive in between, and the second one starts with a determiner, such as 'يَعْقُوبَ و الأَسْبَاطَ', which means Jacob and the Descendants.
4. Proper Noun + Conjunctive "*AND*" + Proper Noun: this rule is for extracting any two Proper nouns with AND conjunctive in between, for example: 'إِبْرَاهِيمَ و إِسْحَاقَ', which means Abraham and Isaac.
5. Proper Noun + Conjunctive "*AND*" + Determiner "ال" + Noun: this rule extracts any Proper noun followed by a noun which starts with a determiner, for example: 'نُوحٍ و النَبِيِينَ', which means Noah and the prophets.
6. Noun + Pronoun + Conjunctive "*AND*" + Noun + Pronoun: this rule extracts any two nouns combined with AND, and the first noun ends with a Pronoun, for example: 'مَحْيَاهُمْ و مَمَاتُهُمْ', which means their life and their death.

Moreover, we define a set of negation conjunctive patterns, where the negation letter 'لا' which means "*NOT*", is used with the conjunctive "*AND*", as clarified next.

7. Negation "*NOT* " + Adjective + Conjunctive "*AND*" + Negation "*NOT* " + Adjective: this pattern finds out any two negative adjectives combined with AND conjunctive, such as 'لَا بَارِدٌ وَ لَا كَرِيمٌ', which means neither cool nor beneficial.
8. Negation "*NOT* " + Determiner "ال" + Noun + Conjunctive "*AND*" + Negation "*NOT* " + Determiner "ال" + Noun: this pattern finds out any two negative nouns combined with AND conjunctive, such as 'لَا الْهَدْيَ وَ لَا الْقَلَائِدَ', which means the sacrificial animals and garlanding.
9. Adjective + Conjunctive "*AND*" + Negation "*NOT*" + Adjective: this pattern finds out any two adjectives combined with AND, where the second one is directly preceded by a negation, such as 'صَغِيرَةٌ و لَا كَبِيرَةٌ', which means small or large.
10. Noun + Conjunctive "*AND*" + Negation "*NOT*" + Noun: this pattern finds out any two nouns combined with AND, where the second one is directly preceded by a negation, such as 'مَالٌ وَ لَا بَنُونَ', which means wealth or children.

4.3. Relation extraction

Different methods have been proposed in the past to find semantic relations between words in a corpus. All of them belong to one of three categories. The first category includes methods which seek finding pair of words that may occur together more often than expected by chance (collocations) using statistical tests (Maedche and Staab, 2001). However, the resulting relations are validated based on approximate decisions achieved by the statistical analyses.

In the second category, researchers have exploited syntactic dependencies, in particular, the dependencies between a verb and its arguments to detect relations. One problem is how to find a general rule to extract the verb arguments related to a specific verb regardless the text where they exist (Cimiano, 2006). The third category methods rely on lexico-syntactic patterns to detect very specific types of relations such as part-of and cause (Hearst, 1992).

The main drawback of these methods is the complexity of pattern construction. It is time and effort consuming since for each type of relations, a set of patterns is developed and applied in a specific form and order.

Our proposed approach is a hybrid of pattern-based methods and statistical methods. However, the pattern-based methods depend on using one/many pattern(s) to extract one specific type of semantic relations. The more types of semantic relations we want to extract, the more patterns we should use. On the other hand, our approach uses a limited set of patterns not to extract a specific type of semantic relations but to extract AND conjunctive phrases from the Quranic Arabic Corpus. Each pattern may extract several types of semantic relations which reduces time and effort complexity. For example, the pattern (Noun + Conjunctive" *AND*" + Noun) extracts three types of relations: Antonymy ('مَوْت وَ حَيَواة'), Gender ('مُؤْمِنِين وَ مُؤْمِنَات'), and Class ('تِين وَ زَيْتُون'), as we will discuss in Section 6.

In order to extract the conjunctive phrases, we search the FORM and the TAG columns in the Quranic Arabic Corpus looking for AND conjunction to extract the two terms that occur on both sides of AND. These phrases that consist of AND conjunction and the two terms are considered as conjunctive phrases only and only if they match one of the patterns defined in the previous section. Each conjunctive phrase indicates the existence of a probable semantic relation which needs further processing, as explained in the next sections.

The tables bellow illustrate some examples of conjunctive phrases, where the two combined terms are adjectives as shown in Table 2, nouns as shown in Table 3, and proper nouns as shown in Table 4. These terms are in the stem form.

Table 2
Sample of conjunctive adjectives.

Adjective 1 AND Adjective 2	English Translation
'صَغِير وَ كَبِير'	Small and Big
'أَعْجَمِىّ وَعَرَبِيّ'	Foreign Tongue and Arab
'بَشِير وَ نَذِير'	Bearer of glad tidings and Warner
'أَعْمَى وَ بَصِير'	Blind and Seeing
'لَا فَارِض وَ لَا بِكْر'	Neither old nor young
'سَاجِد وَ قَائِم'	Prostrating and Standing

Table 3
Sample of conjunctive nouns.

Noun 1 AND Noun 2	English Translation
'جَنّٰة وَ حَبّ'	Garden and Grain
'جَنّٰة وَ حَرِير'	Garden and Silk
'جَنّٰة وَعَيْن'	Garden and Spring
'جَنّٰة وَ مَغْفِرَة'	Garden and Forgiveness
'جَنّٰة وَ نَعِيم'	Garden and Bliss
'جَنّٰة وَ نَهَر'	Garden and Stream

Table 4
Sample of conjunctive proper nouns.

Proper Noun 1 AND Proper Noun 2	English Translation
'يَأْجُوج وَ مَأْجُوج'	Gog and Magog
'جِبْرِيل وَ مِيكال'	Gabriel and Michael
'مُوسَى وَ هَارُون'	Moses and Aaron
'إِنجِيل وَ قُرْءَان'	Gospel and Qur'an
'فِرْعَوْن وَ هَامَان'	Pharaoh and Haman
'إِسْحَاق وَ يَعْقُوب'	Isaac and Jacob
'دَاوُد وَ سُلَيْمَان'	David and Solomon
'عَاد وَ ثَمُود'	'Ad and Thamud

Moreover, we discover a special case of combinations where one term is associated with many different terms, as they have occurred in the holy Quran. For example, the term Garden 'جَنّٰة' is combined with six different terms as illustrated in Table 3.

4.4. Correlation coefficient

One main feature of AND conjunction is that the two combined terms must be in a kind of correlation with each other (AL-Taweel, 2009). There are several well-known correlation techniques that are reported in the literature. Mutual Information (MI) is a common approach that measures the strength of the association between a pair of variables based on the number of times the pair occurs together versus the number of times the pair occurs separately (Dunning, 1993).

$$MI = \log_2 \frac{f_{AB}}{f_A f_B} \tag{2}$$

where f_A and f_B are the number of occurrences of word A and word B in the whole corpus and f_{AB} is the number of occurrences of the two words together in the corpus.

Another method is called t-score and used to measure the confidence with which an association between a pair of variables can be asserted (de Winter, 2013).

$$\text{T } score = \frac{f_{AB} - \frac{f_A f_B}{N}}{\sqrt{f_{AB}}} \tag{3}$$

These two techniques are highly dependent on the frequency of variables which provides an unbounded score. This leads to the ambiguity in interpreting the resulting association score. Moreover, no extra information about the association relationship except its strength can be predicted. As a validation step, no further statistical tests can be applied to the result to measure its significance. Because of these limitations, we decide to use a more powerful association measure called Pearson Product-Moment Correlation Coefficient (r) to find how much two terms are related to each other (Farreús et al., 2012; Ngan, 2013). This method outperforms the previous ones since it allows researchers to investigate naturally the relation between any two variables and interpret the results clearly without any misleadingly incorrect values since they lie between -1 and +1. The strength of the relationship as well as its direction can be predicted very distinctly from the obtained results. More efficiently, it affords several asymptotic statistical tests for calculating the significance levels of the scores and produces more accurate and reliable results. This advantage may be of a great value for users to feel confident that such sacred text is treated carefully.

The correlation coefficient has many other variations such as Spearman Rank Correlation Coefficient and Partial correlation. However, each type is applied to specific types of variables and under specific circumstances. As an example, Spearman Rank Correlation Coefficient has been advised when the data is in terms of

Table 5
Comparison between (r), (MI), and t-score association measures of different conjunctive phrases.

The conjunctive phrase	English Translation	Correlation Coefficient *(r)*	Mutual Information (MI)	t-score
‘رَعْد وَ بَرْق’	Thunder and Lightning	0.6115	15.117	1.000
‘سَمَاء وَ أَرْض’	Heaven and Earth	0.8485	7.464	11.596
‘ظُلُمَات وَ رَعْد’	Darkness and Thunder	0.2068	13.532	1.000
‘ذِلَّة وَ مَسْكَنَة’	Abasement and Destitution	0.1325	14.702	1.000
‘هُزُو وَ لَعِب’	Jest and Sport	0.0139	12.658	1.414
‘هُدَى وَ نُور’	Guidance and Light	−0.0054	7.571	1.407

ranks since it tries to assess the relationship between ranks without making any assumptions about the nature of their relationship. In our work, we select Pearson Product-Moment correlation coefficient *(r)* which is best suited to the nature of our data and the results we want to achieve.

In statistics, r is defined as a measure of the degree of linear relationship between two variables A and B (Myatt, 2007).

$$r = \frac{\sum_m \sum_n (\boldsymbol{A}_{mn} - \overline{A})(\boldsymbol{B}_{mn} - \overline{B})}{\sqrt{\left(\sum_m \sum_n (\boldsymbol{A}_{mn} - \overline{A})^2\right)\left(\sum_m \sum_n (\boldsymbol{B}_{mn} - \overline{B})^2\right)}} \tag{4}$$

where $\overline{\text{A}}$ and $\overline{\text{B}}$ are the means of A and B elements, respectively, m and n are their sizes. The value of r ranges between −1 and +1. The sign of the correlation coefficient (+, −) defines the direction of the relationship, either positive or negative, whereas the absolute value of the correlation coefficient measures the strength of the relationship. Thus, we apply this method to each conjunctive phrase extracted in Section 4.3. In this case, the two variables A and B are the two combined terms, and their elements are the tf.idf weights, found in the term-document matrix. As a result, we find either a positive correlation coefficient, which means that as the weight of one term increases, the weight of the other term increases; as one decreases the other decreases or a negative correlation coefficient, which indicates that as one term's weight increases, the other decreases, and vice-versa. The values of −1 and +1 mean a perfect linear relationship between the two terms, while the zero value indicates the absence of this type of relation.

In Table 5, we have applied the above three association measures to a set of conjunctive phrases. As an example, consider the phrase *"Heaven and Earth"*, the term *"Heaven"* occurs 185 times and the term *"Earth"* occurs 444 times in the Quranic Arabic Corpus whereas they co-occur 133 times. The MI is not particularly high (7.464) because both terms *"Heaven"* and *"Earth"* are high-frequency words in the Quran corpus. The t-score however is quite high (11.596) since it has taken into account the actual number of observations of the pairing to estimate the confidence value in claiming an association. According to the correlation coefficient, the value (0.8485) indicates that there is a strong positive relationship between *"Heaven"* and *"Earth"* where we can predict that as the weights of one term increase, the other term's weights increase too.

Regardless the drawbacks and the different ways these methods work in, they are very useful in measuring the correlation value between variables, and researchers can select the best one suited to their applications.

Table 6 demonstrates a sample of conjunctive phrases with high positive correlation. Term 1 and Term 2 may share a strong relationship. For example, the two terms Heaven and Earth have a high correlation because they often appear together as a conjunctive phrase in the Quranic verses.

Table 7 demonstrates a sample of conjunctive phrases with low positive correlation. Term 1 and Term 2 may share a weak relationship. This is due to the low occurrence percentage of the two terms together compared to their independent occurrence. The term Allah has a weak relation with the term day since we find the term Allah almost in every verse in the Quran and this is not the case for the term day.

Table 6
Sample of conjunctive phrases with high correlation.

Term 1 AND Term 2	English Translation	Correlation Coefficient
‘رَعْد وَ بَرْق’	Thunder and Lightning	0.6115
‘سَمَاء وَ أَرْض’	Heaven and Earth	0.8485
‘شَاهِد وَ مَشْهُود’	Witness and Whom witness has been borne	0.7698

Table 7
Sample of conjunctive phrases with low correlation.

Term 1 AND Term 2	English Translation	Correlation Coefficient
‘ظُلُمَات وَ رَعْد’	Darkness and Thunder	0.2068
‘ذِلَّة وَ مَسْكَنَة’	Abasement and Destitution	0.1325
‘الله وَ يَوْم’	Allah and Day	0.1678

Table 8
Sample of conjunctive phrases with close to zero correlation.

Term 1 AND Term 2	English Translation	Correlation Coefficient
‘وَجْه وَ يَد’	Face and Hand	0.0844
‘هُدًى وَ نُور’	Guidance and Light	-0.0054
‘هُزُو وَ لَعِب’	Jest and Sport	0.0139

Table 9
Sample of conjunctive phrases with correlation equal to 1.

Term 1 AND Term 2	English Translation	Correlation Coefficient
‘شَفْع وَ وَتْر’	Even and Odd	1
‘شتَاء وَ صَيْف’	Winter and Summer	1
‘جَلَال وَ إِكْرَام’	Glory and Honor	1

Table 8 demonstrates a sample of conjunctive phrases with close to zero correlation. Term 1 and Term 2 may share no relationship. This set of combined terms may appear together very rarely. To the contrary, each term may appear alone or combined with a different term many times. As an example, the two terms guidance and light may have no relation because the term guidance is also associated with many other terms such as good tidings ‘بُشْرَى’, reminder ‘ذِكْرَى’, mercy ‘رَحْمَة’, healing ‘شِفَاء’, criterion ‘فُرْقَان’, and instruction ‘مَوْعِظَة’.

In addition, there is a perfect set of conjunctive phrases where both terms occur only together equal number of times. As a result, their correlation coefficient is 1 as shown in Table 9.

5. Validation phase

Texts in general and Quranic Arabic texts in particular can be understood by scholars from different aspects. This reason leads

Table 10
Sample of accepted and rejected conjunctive relations R after applying t- test.

Term 1	AND	Term 2	Correlation Coefficient *(r)*	t-value	*t*-test decision
جِنَّة Jinn	و	نَاس Men	0.5711	7.3628	R may be accepted
فَاكِهَة Fruits	و	أَب Herbage	0.4993	6.0987	R may be accepted
ضّحَى Morn	و	لَيْل Night	0.4006	4.6271	R may be accepted
سَمَاء Heaven	و	طَارِق Morning Star	0.0768	0.8152	R may be rejected
كَذَّاب Liar	و	كُل Every	-0.0323	−0.3420	R may be rejected
الله Allah	و	فَتْح Victory	0.0627	0.6649	R may be rejected

to the extraction of different linguistic and religious patterns. In order to assess such results, we find that adopting statistical techniques is very useful to support the best possible decision about problems related to text processing.

In our work, because the weight of each term is based on its frequency in the corpus, the correlation coefficient is then highly dependent on this factor. One approach to ensure that two terms are together because of a type of relationship and not due to chance is to use statistical hypotheses testing (Kass et al., 2014) and test the significance of the correlation coefficient.

We use two mutually exclusive hypotheses called null hypothesis H_0 and alternative hypothesis H_1.

H_0: There is no correlation between the two terms and their co-occurrence is due to chance.
H_1: There is a significant correlation between the two terms.

Next, we test the two hypotheses to either reject the null hypothesis or accept it by applying a statistical test named the Student's t test (Siegmund, 1998).

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} \tag{5}$$

It returns a value t which shows the validity of null hypothesis. The smaller the t-value, the weaker is the evidence against the null hypothesis. Then, we compare the t-value to an acceptable significance threshold $a = 1.984$, taken from statistical t-test tables (Verma, 2013). The tabulated value of a is required for significance at .05 level and n -2 degree of freedom, where n is equal to 114, which is the size of each term vector, and r is the calculated correlation coefficient obtained from formula (4). If $t > a$, the correlation coefficient is statistically significant, the null hypothesis may be rejected and the alternative hypothesis is valid. Otherwise, if $t \leq a$, the null hypothesis is true. Table 10 illustrates a sample of probable accepted and rejected conjunctive relations, after applying t-test.

6. Experimental results

We use the Quranic Arabic Corpus which consists of 77,430 words. We start with words that may represent the Quranic domain such as nouns, proper nouns, and adjectives. Basically, we find a set of 31007 repeated words, filtered to 3267 unique terms.

Besides that, conjunctive phrases occurred in the holy Quran almost 2000 times. After eliminating the repeated ones, we get a set of 1047 unique phrases and hence probable relations.

Because automatically learned ontologies are highly error prone, there is an immense need for domain-specific experts to inspect, validate, and modify them before they can be applied. We may suggest a filtering method to select the most representative relations by defining a threshold and select the phrases with correlation coefficient greater than this threshold. Although we find that the statistical t-test is very efficient in the filtering process, the threshold- based method can also be used as a next step.

Compared to the previous approaches mentioned in Section 4.3, our novel method is a hybrid between the statistical approach and the pattern-based approach. However, it is based on a strong Arabic grammatical construction, namely the AND conjunction, to define a small set of patterns allowing the extraction of many types of semantic relations from Quran very efficiently. Furthermore, we exploit statistical methods to measure the strength and the direction of the extracted relations and to aid domain experts in assessing the final semantic relations. Classifying each relation as antonymy, gender, or class is performed manually with a view to doing it automatically in the future. The proposed approach achieved more accurate and comprehensive results.

The extracted relations are classified manually into three categories: Antonymy, Gender, and Class.

6.1. Antonymy

Antonymy is the semantic relation between antonyms which are words with opposite meaning. In Table 11, this category of relations includes antonyms that are combined by AND conjunction. For example, it was found the term Sky 'سَمَاء' is combined with its antonym Land 'أَرْض' and the term Secretly 'سِر' is combined with two different antonyms: Openly ' جَهْرّ' and Openly ' عَلاَنِيَة'.

6.2. Gender

This category refers to the relation that exists between masculine and feminine words. Such relation is very common in the holy Quran where God converses with both males and females at once. Table 12 lists a sample of conjunctive phrases that combine masculine and feminine terms together such as the masculine term charitable men 'مُصّدقِين' and its feminine counterpart charitable women ' مُصّدقَات'.

6.3. Class

A different category of semantic relations consists of terms that belong to the same class because they share the same characteristic features. The holy Quran is full of such examples that are combined by AND conjunction. Table 13 mentioned few of them such as the terms seven and eighth 'سَبْعَة وَ ثَامِن' which belong to the class Numeral. Also, the class Book includes the terms

Table 11
Sample of conjunctive phrases with Antonymy relation.

Term 1 AND Term 2	English Translation
'سَمَاء وَ أَرْض'	Sky and Land
'سِر وَ جَهْرّ'	Secretly and Openly
'سِرّ وَ عَلاَنِيَة'	Secretly and Openly
'شَرّ وَ خَيْر'	Evil and Good
'شَقِيّ وَ سَعِيد'	Unhappy and Happy
'شتَاء وَ صَيْف'	Winter and Summer
'ضَرّاء وَ سَرّاء'	Adversity and Joy
'لَيْل وَ نَهَار'	Night and Day
'مَشْرِق وَ مَغْرِب'	East and west
'مَوْت وَ حَيَواة'	Death and Life

Table 12
Sample of conjunctive phrases with Gender relation.

Term 1 AND Term 2	English Translation
'مُؤْمِنِين وَ مّؤْمِنَات'	The believing men and believing women
'مُتَصَدقِين وَ مُتَصَدقَات'	The charitable men and charitable women
'مُسْلِمِين وَ مُسْلِمَات'	The Muslim men and Muslim women
'مُصٰدقِين وَ مُصٰدقَات'	The charitable men and charitable women
'مُنَافِقُون وَ مُنَافِقَات'	The hypocrite men and hypocrite women
'صَابِرِين وَ صٰابِرَات'	The patient men and patient women
'صَادِقِين وَ صٰادِقَات'	The truthful men and truthful women
'صٰائِمِين وَ صٰائِمَات'	The fasting men and fasting women
'قَانِتِين وَ قَانِتَات'	The obedient men and obedient women
'بَنِين وَ بَنَات'	Sons and daughters

Table 13
Sample of conjunctive phrases with Class relation.

Term 1 AND Term 2	English Translation
'آللٰات وَ آلْعُزٰى'	Lat and Uzza
'أَبَارِيق وَ كَأْس'	Ewers and Cups
'أَصْوَاف وَ أَوْبَار'	Wool and Furs
'أَنف وَ أُذُنٌ'	Nose and Ears
'أَيّوب وَ يُوسُف'	Job and Joseph
'إِنجِيل وَ قُرْءَان'	Gospel and Qur'an
'بَقَر وَ غَنَم'	Cow and Sheep
'تين وَ زَيْتُون'	Fig. and Olive
'ذَهَب وَ فِضٰة'	Gold and Silver
'سَبْعَة وَ ثَامِن'	Seven and Eighth

Gospel and Quran 'إِنجِيل وَ قُرْءَان', and the class Animal includes Cow and Sheep 'بَقَر وَ غَنَم'.

It is clear in our novel approach that we have used only one type of patterns, namely conjunctive, to extract different types of semantic relations. In order to categorize them, we could train classifiers for each type of relations and combine their results and test on different types of extracted ontological relations.

7. Evaluation

To evaluate the accuracy of the relation extraction process, we have used the two performance metrics: precision and recall. In our work, precision is defined as the ratio of the number of relevant retrieved conjunctive relations to the number of retrieved conjunctive relations whether relevant or not, whereas the recall is defined as the ratio of relevant retrieved conjunctive relations to the total number of all relevant conjunctive relations that exist in Quran. The system retrieves 1047 semantic relations based on the predefined conjunctive patterns, 57% are statistically classified as strong relations and 43% are classified as weak relations.

Furthermore, the extracted relations are validated manually by domain experts. The system achieves a precision of 84% and a recall of 92%. More details about the evaluation results are described in Table 14.

We can notice that our method has very high precision and recall since a total of 878 relations are correctly retrieved by the system from a set of 950 relevant relations. On the other hand, the system extracts some erroneous relations of irrelevant cases because of two main reasons:

Table 14
The evaluation results of the relation extraction method.

Total number of retrieved relations by the system	1047
Total number of relevant relations retrieved by the system and validated by domain experts	878
Total number of all relevant relations that exist in the Quran	950
Precision	84%
Recall	92%

(34:13:12:1) اعْمَل V
STEM|POS:V|IMPV|LEM:Eamila|ROOT:عمل|2MP
(34:13:12:2) وا PRON
SUFFIX|PRON:2MP
(34:13:13:1) آلَ N
STEM|POS:N|LEM:'aAl|ROOT:أول|M|ACC
(34:13:14:1) دَاوُدَ PN
STEM|POS:PN|LEM:داوود|GEN
(34:13:15:1) شُكْرًا N
STEM|POS:N|LEM:شكر|ROOT:$kr|M|INDEF|ACC
(34:13:16:1) وَ CONJ
PREFIX|w:CONJ+
(34:13:16:2) قَلِيلٌ N
STEM|POS:N|LEM:qaliyl|ROOT:قلل|MS|INDEF|NOM
(34:13:17:1) مِنْ P
STEM|POS:P|LEM:min
(34:13:18:1) عِبَادِ N
STEM|POS:N|LEM:Eabod|ROOT:عبد|MP|GEN
(34:13:18:2) يَ PRON
SUFFIX|PRON:1S
(34:13:19:1) اَلْ DET
PREFIX|Al+
(34:13:19:2) شَكُورُ ADJ
STEM|POS:ADJ|LEM:$akuwr|ROOT:شكر|MS|NOM

Fig. 4. Example of an erroneous annotation in the Quranic Arabic Corpus.

- Mistakes in the annotation of the Quranic Arabic Corpus specifically in the Arabic traditional grammar (الإعراب). Although this corpus has achieved an accuracy rate of 99%, these errors still exist due to the difficulty of the Arabic language and the lack of efficient validation methods. One example of errors appears in the Fig. 4 in the verse 13 of Saba' chapter ('اعملوا آل داوود شكرا و قليل من عبادي الشكور') which means *("Work, O family of David, in gratitude." And few of My servants are grateful)*. The tool AND (shown in bold in the corpus) is wrongly annotated as a conjunction while it is a circumstantial particle.

(16:27:16:1) إِنَّ ACC
STEM|POS:ACC|LEM:إن|SP:<in~
(16:27:17:1) الْ DET
PREFIX|Al+
(16:27:17:2) خِزْيَ N
STEM|POS:N|LEM:خزي|ROOT:xzy|M|ACC
(16:27:18:1) الْ DET
PREFIX|Al+
(16:27:18:2) يَوْمَ N
STEM|POS:N|LEM:يوم|ROOT:ywm|M|ACC
(16:27:19:1) وَ CONJ
PREFIX|w:CONJ+
(16:27:19:2) الْ DET
PREFIX|Al+
(16:27:19:3) سُوءَ N
STEM|POS:N|LEM:سوء|ROOT:swA|M|ACC
(16:27:20:1) عَلَى P
STEM|POS:P|LEM:على
(16:27:21:1) الْ DET
PREFIX|Al+
(16:27:21:2) الكَافِرِينَ N
STEM|POS:N|ACT|PCPL|LEM:كافرون|ROOT:kfr|MP|GEN

Fig. 5. Example of an erroneous conjunctive phrase.

- Lack of grammatical details in the annotated Quranic corpus especially those related to the conjunctive phrase. Although the annotation specifies the type of the tool AND whether it has a conjunctive role or not and the type of the two terms combined by AND, it does not provide much information about their grammatical positions in the conjunctive phrase (المعطوف و المعطوف عليه). This problem leads to extracting wrong conjunctive phrases and hence irrelevant semantic relations. Fig. 5 shows the verse 27 of An-Nahl chapter ('إن الخزي اليوم و السوء على الكافرين') which means *("Indeed disgrace, this Day, and evil are upon the disbelievers*"). The erroneous conjunctive phrase (day and evil) which means 'اليوم و السوء' is retrieved instead of the correct one (disgrace and evil) which means 'الخزي و السوء'.

The precision of the system could be increased by using more annotation details in the corpus related to the Arabic traditional grammar (إعراب) and the aid of domain experts.

8. Conclusion

Quran ontologies aim at providing a comprehensive knowledge about Quran and improving the performance of information retrieval systems that rely on Quranic texts. However, there is a lack in the ontologies that deal with Quranic Arabic due to the complex morphology structure of the Arabic and the shortage in tools and resources that support it. Currently, the existing ontologies are limited to cover only parts of the Quran or specific types of words or topics. In this paper, we have exploited the Arabic conjunctive patterns that exist in the traditional Arabic grammar to extract different types of semantic relations from the entire Quran and enrich the automatic construction of the Quran ontology. We have applied a correlation coefficient method to measure the strength of the linear relationship which may exist between every pair of nouns, proper nouns, and adjectives that form a conjunctive phrase. Furthermore, we have suggested hypothesis testing and Student t-test to go beyond chance and validate the significance of the extracted relations. We have unveiled manually three categories of semantic relations: antonymy, gender, and class. In future work, we can exploit classifiers to perform this task automatically. Finally, we insist that such a field of research needs statistical techniques besides domain experts to assess the results achieved.

References

Abbas, N., 2009. Qurany: A Tool to Search for Concepts in the Quran (MSc Research Thesis). School of Computing. University of Leeds, Leeds, UK.

Abdelnasser, H., Mohamed, R., Ragab, M., Mohamed, A., Farouk, B., El-Makky, N., Torki, M., 2014. Al-Bayan: an Arabic question answering system for the Holy Quran. In: Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP). Doha, Qatar, pp. 57–64.

Adhima, M., 1972. Derasat li Osloob Al-Quraan Al-Karim. In: Arabic. Cairo, Egypt: Dar Al-Hadith.

Al-Ghalayini, M., 2007. Jamea Al-Dorous Al-Arabiya. In: Arabic. Cairo, Egypt, Dar Al-ghad Al Jadid.

Alrehaili, S.M., Atwell, E., 2014. Computational ontologies for semantic tagging of the Quran: a survey of past approaches. In: LREC 2014 Proceedings, European Language Resources Association.

Al-Taweel, M., 2009. Significance of conjunctions and the impact thereby on differences Muslim Scholars (Foqaha). In: Arabic. Master Dessertation. An-Najah National University. Nablus. Palestine.

Alvarez, F., Vaquero, A., Sáenz, F., De Buenaga, M., Flores, Á., Palacios, V.R., 2007. Semantic Relation Modeling and Representation for Problem-Solving Ontology-Based Linguistic Resources: Issues and Proposals. In Proceedings of the 4th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2007), ISBN 978-972-8865-97-9, INSTICC Press, pp. 59–70.

Al-Yahya, M., Al-Khalifa, H., Bahanshal, A., Al-Odah, I., Al-Helwah, N., 2011. An ontological model for representing semantic lexicons: an application on time nouns in the holy Quran. 2C. Arab. J. Sci. Eng. 35 (2), 21–35.

Al-Zujaji, A., 1984. Horouf Al-Maani. In: Arabic. Beirut, Lebanon, Muassasat Al-Ressala.

Baqai, S., Basharat, A., Khalid, H., Hassan, A., Zafar, S., 2009. Leveraging Semantic Web Technologies for Standardized Knowledge Modeling and Retrieval from the Holy Qur'an and Religious Texts. In Proceedings of the 7th International Conference on Frontiers of Information Technology (p. 42). Pakistan.

Bloehdorn, S., Hotho, A., 2004. Text classification by boosting weak learners based on terms and concepts. In Proceedings of the 4th IEEE International Conference on Data Mining (ICDM), pp. 331–334.

Cimiano, P., 2006. Ontology Learning and Population from Text: Algorithms, Evaluation and Application. Springer, New York, USA.

de Winter, J.C., 2013. Using the Student's t-test with extremely small sample sizes. Pract. Assess. Res. Eval. 18 (10), 1–12.

Dukes, K., Atwell, E., 2012. LAMP: a multimodal web platform for collaborative linguistic analysis. In Proceedings of the Language Resources and Evaluation Conference (LREC), Turkey, pp. 3268–3275.

Dukes, K., Habash, N., 2010. Morphological annotation of Quranic Arabic. In: The 7th International Conference on Language Resources and Evaluation (LREC), Valletta, Malta, pp. 2530–2536.

Dunning, T., 1993. Accurate methods for the statistics of surprise and coincidence. Comput. Linguistics 19 (1), 61–74.

Elabd, E., Alshari, E., Abdulkader, H., 2015. Semantic boolean Arabic information retrieval. Int. Arab. J. Inf. Technol. (IAJIT) 12 (3), 311–316.

Farghaly, A., Shaalan, K., 2009. Arabic natural language processing: challenges and solutions. ACM Trans. Asian Lang. Inf. Process. (TALIP) 8 (4), 1–22.

Farreús, M., Costa-jussà, M.R., Morse, M.P., 2012. Study and correlation analysis of linguistic, perceptual, and automatic machine translation evaluations. J. Assoc. Inf. Sci. Technol. 63 (1), 174–184.

Gruber, T., 1993. A Translation Approach to Portable Ontology Specification, 5, 199-220.

Guarino, N., Vetere, G., Masolo, C., 1999. Ontoseek: content-based access to the web. IEEE Intell. Syst. 14 (3), 70–80.

Habash, N., 2010. Introduction to Arabic Natural Language Processing. Morgan & Claypool Publisher, California, USA.

Hakkoum, A., Raghay, S., 2016. Semantic Q&A system on the Qur'an. Arab. J. Sci. Eng. 41 (12), 5205–5214.

Hearst, M.A., 1992. Automatic acquisition of hyponyms from large text corpora. In Proceedings of the 14th Conference on Computational linguistics, 2, 539–545, Nantes, France.

Ibn Kathir, I., 1999. Tafsir Ibn Kathir. In: Arabic. Riyadh, Saudi Arabia. Dar Tayba.

Kass, R.E., Eden, U.T., Brown, E.N., 2014. Analysis of Neural Data. Springer, New York, USA.

Lee, C.S., Kao, Y.F., Kuo, Y.H., Wang, M.H., 2007. Automated ontology construction for unstructured text documents. Data Knowl. Eng. 60 (3), 547–566.

Liu, K., Hogan, W.R., Crowley, R.S., 2011. Natural language processing methods and systems for biomedical ontology learning. J. Biomed. Inform. 44 (1), 163–179.

Maedche, A., Staab, S., 2001. Ontology learning for the semantic web. IEEE Intell. Syst. 16 (2), 72–79.

Myatt, J.G., 2007. Making Sense of Data, a Practical Guide to Exploratory Data Analysis and Data Mining. John Wiley & Sons, New Jersey, USA.

Navigli, R., Velardi, P., Gangemi, A., 2003. Ontology learning and its application to automated terminology translation. IEEE Intell. Syst. 18 (1), 22–31.

Ngan, S.C., 2013. Correlation coefficient of linguistic variables and its applications to quantifying relations in imprecise management data. Eng. Appl. Artif. Intell. 26 (1), 347–356.

Nirenburg, S., Raskin, V., 2004. Ontological Semantics. Mit Press.

Saad, S., Salim, N., Zainal, H., Noah, S.A.M., 2010. A framework for Islamic knowledge via ontology representation. In: 2010 International Conference on Information Retrieval & Knowledge Management (CAMP). IEEE, pp. 310–314.

Salton, G., McGill, M.J., 1986. Introduction to Modern Information Retrieval. McGraw-Hill, New York, USA.

Sharaf, A., Atwell, E., 2012. QurAna: corpus of the Quran annotated with pronominal Anaphora. In: Language Resources and Evaluation Conference (LREC 2012), Istanbul, Turkey, pp. 130–137.

Shoaib, M., Yasin, M.N., Hikmat, U.K., Saeed, M.I., Khiyal, M.S.H., 2009. Relational WordNet model for semantic search in Quran. In: 2009 International Conference on Emerging Technologies. IEEE, pp. 29–34.

Siegmund, B., 1998. Data Analysis, Statistical and Computational Methods for Scientists and Engineers. Springer, New York, USA.

Ta'a, A., Abidin, S.Z., Abdullah, M.S., Ali, B., Ahmad, M., 2013. Al-Quran themes classification using ontology. In: Proceedings of the 4th International Conference on Computing and Informatics ICOCI 2013, Malaysia, pp. 383–389.

Ul Ain, Q., Basharat, A., 2011. Ontology driven Information Extraction from the Holy Quran related Documents. In: The 26th Institution of Electrical and Electronics Engineers Pakistan Students' Seminar (IEEEP 2011), Karachi, Pakistan.

Verma, J.P., 2013. Data Analysis in management with SPSS Software. Springer, New Delhi, India.

Weiss, S.M., Indurkhya, N., Zhang, T., Damerau, F.J., 2005. Text Mining Predictive Methods for Analyzing Unstructured Information. Springer, New York, USA.

Yauri, A.R., Kadir, R.A., Azman, A., Murad, M.A.A., 2012. Quranic-based concepts: verse relations extraction using manchester OWL syntax. In: International Conference on Information Retrieval & Knowledge Management (CAMP). IEEE, pp. 317–321.