Semantic Web 0 (0) 1
IOS Press

An Ontology for the Tajweed of the Quran

Amna Basharat [a,*], Ramsha Amin [a]
[a] *Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan*
E-mail: amna.basharat@nu.edu.pk, i171041@nu.edu.pk

Abstract. In the current information systems, many fields use ontologies for modeling domain knowledge to enable interoperable semantics. There is a plethora of knowledge sources within the Islamic heritage that derive from the primary sources of the Quran and the Hadith (Prophetic narrations), however, there is lack of sufficient ontologies and linked data to better describe and semantically annotate the information related to the Quran. Although several Quranic themes based ontologies have been developed to facilitate the retrieval of knowledge from the Quran, there is still a lack of comprehensive knowledge-based reasoning models created for the Tajweed of the Quran - the science of Quranic recitation. In this paper we propose the design of an ontology for capturing the core elements of Quranic recitation i.e. Tajweed. The knowledge model was developed by using the Protege framework and state-of-the-art semantic web technologies (OWL and SPARQL). METHONTOLOGY, an iterative design methodology was used for its development. The ontology focuses on describing the articulation points of Arabic letters and their characteristics together with the Tajweed rules (rules of recitation). Semantic Web Rule Language (SWRL) was used for the implementation of the Tajweed rules. To evaluate the ontology model, a hybrid approach was used. Expert driven validation and criteria based evaluation was conducted for the Arabic letters and their characteristics to evaluate the accuracy and structure of ontology. Results from the experts were incrementally improved before evaluating it with the next expert which results in 100% accuracy. Tajweed rules were evaluated using data driven approach on the complete text of the Holy Quran. Also, an annotated dataset of the entire Quran was generated in OWL format using the developed Tajweed ontology.

Keywords: Domain Ontology, Tajweed,, Quran,, Hybrid Evaluation,, Semantic Web

1. Introduction

Over the past few years, emerging technologies have profoundly revolutionized the different ways of interacting with knowledge and information. Currently, ontologies are used for modeling domain knowledge in different fields as the solution for semantic interoperability [1]. The integration of Islamic knowledge resources on a web-scale so far is not adequate. Many researchers have initiated to exploit ontologies to improve the interoperable access to religious knowledge resources such as the Quran and the Hadith (Prophetic Narrations) [2]. Creating ontologies in the field of Islamic knowledge will make easier for the intelligent systems to process and understand data automatically and will help in providing its users with improved knowledge retrieval.

The Quran is the primary source of knowledge and guidance for Muslims around the world. Literal meaning of the Quran is "recitation". The language of the Quran is Arabic and each Arabic letter in the Quran has an articulation point from which it originates and have some characteristics. Tajweed is the art and science of the recitation of the Quran. The current Quranic domain retrieval methods lack adequate semantic search capabilities; they are mostly based on the matching keywords approach. Therefore, this is an emerging area of research since more evidence on the accuracy and completeness of the development of Tajweed ontology is needed which is crucial for the successful development of the ontology.

*Corresponding author: amna.basharat@nu.edu.pk

By using semantic technologies we designed and implemented a multilingual ontology based Quranic Tajweed knowledge model consisting of the following core elements: (i) Articulation points of the Arabic letters (ii) Arabic letters' characteristics (iii) Arabic Letters' grouping (iv) Letter occurrences and Rule occurrences for the Quranic text and (v) Tajweed rules. In addition, a complete annotated Quranic Tajweed dataset has been constructed of high accuracy in the OWL format which captures semantic annotations using the developed Tajweed ontology. Furthermore we performed formal evaluation of the Tajweed ontology knowledge model by using hybrid approach i.e, criteria based, expert based and data driven approach.

The remainder of the paper will include: (a) a brief over-view of the literature review work in two areas, (1) ontology engineering methodologies and (2) the creation of ontologies in the Quranic domain, in particular in the Tajweed domain; (b) a detailed explanation of the key components of Tajweed ontology; (c) description of the design and implementation of the annotated Quran dataset using the Tajweed Ontology, with an example case study, (d) validation of the Tajweed ontology using a variety of validation approaches; and (e) concluding reflections on the potential benefits and the limitations of the Tajweed ontology.

2. Background and Related Work

In this section, brief information about the literature related to our work is presented. We begin with a short review of the literature regarding different methodologies and evaluation approaches that have been used by the researchers for the ontology development. We also review the existing work in the area of Quranic Tajweed, which primarily revolves around automated assessment approaches that have been used in the area of recitation of the Quran. Finally, we state the differences between the related work and our approach in the discussion section.

2.1. Ontology Engineering Methodologies

Semantic technologies, in particular ontology engineering, provide a systematic definition and a semantically rich knowledge base for defining and understanding the words in a domain and their relationships between them [6]. Several methods and methodologies were proposed as a guide for ontology development. Ontology development 101 [3] suggests a seven-step method for the development of the ontology using Protégé editor [4]. The approach introduces competency questions and reuse of ontology as well as practical advice on the development of the ontology. METHONTOLOGY [5] is another popular ontology design methodology. It is the most mature design methodology developed using IEEE 1074-1995 standard for the software development phase as a foundation. It involves the life cycle of ontology development and the techniques for achieving each activity from the specification to the maintenance of the implemented ontology. The On-To-Knowledge methodology focuses on application-based ontology development [6]. Engineering and industrial experts are actively involved in the initial stages of the development. The eXtreme Design (XD) methodology reflects an agile approach to ontology development that emphasizes the collaborative and incremental development of ontology using ontology design patterns(odp) [7]. Another methodology NeOn [8] focuses on nine scenarios that adopt a different vision from current methodologies. It does not identify workflows for the ontological development process, but a collaborative approach for the development of ontology emphasizing on reusing and reengineering ontological and non-ontological knowledge resources.

Since the development of an ontology is an iterative approach, Noy et al. [3] stated that there is no one way of designing and developing the ontology correctly, the best solution depends on the application you are considering and the extensions you anticipate.

We discovered through different methodology studies that, most of the latest ontology design methodologies follow the same life cycle i.e. specification, conceptualization, implementation, and evaluation. We selected METHONTOLOGY methodology for developing our ontology.

2.2. Existing Work on Quranic Recitation

Numerous research studies have been reviewed that focus on classifying the automated recognition system for correct pronunciation in the area of the Quranic Tajweed. Different classification algorithms such as Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) [9] were used to evaluate the accuracy performance in detecting the Quranic letters [9]. Speech recognition process needs the conversion of audio signals which are identified by the audio feature extraction such as Mel-frequency cepstral coefficients (MFCC)[9],[10],[11],[12] and Linear Predictive Coding (LPC)[13] and classify the speech using some machines learning and deep learning techniques such as Support Vector Machines (SVM) [12],[14] Hidden Markov Model (HMM)[15][10][16], K-Nearest Neighbors (KNN)[17],[11],[18], Bidirectional Long Short-Term Memory (BLSTM)[11],[19]. Natural language processing (NLP) techniques are widely being used to extract the information from the text. In the domain of Islamic Knowledge, a good amount of work has been previously proposed where conceptual features extracted from ontologies were used to enhance the bag of words model [20]. A semantically enhanced search tool [21], [22] uses Quranic concepts and develop the ontologies through manual and semi-automatic techniques and then merge both these ontologies into one Quran ontology which produces more accurate results. Semantically answering questions system were developed [23] using the ontology in which knowledge extraction was carried out through NLP approach. Some Rule-based system by Alfaries et al. [24] proposed the QurTaj application by using the NLP technique to extract four Tajweed rules i.e. Noon sakenah, Tanween, Meem sakenah, and Madd.

From literature we identified that most of the work in the area of Tajweed is based on automated assessment using computational techniques (Machine learning, deep learning techniques, etc). Using these methods is beneficial in the detection of the wrong pronunciation of the rules but existing methods do not provide formalized and standard knowledge models in the Islamic domain. In addition, lack of suitable training data and gold standards add issues like reliability and scalability. Moreover, as mentioned by Silva et al. [25], these statistical techniques alone are not sufficient for the classification of data. Better classification can be achieved by considering the semantics of that data which will result in more classification accuracy. Additionally, Bloehdorn et al. [26] suggested that ontologies can help in smarter information retrieval and support better and faster parsing and can provide potential benefits for a lot of applications such as text classification and clustering.

2.3. Review of Ontologies in the Quranic Domain

Over the last few years, many research studies have been introduced in the area of the ontology development for the Islamic knowledge domain. The existing literature shows a classification of Quran related ontology models into two categories: Quranic theme-based ontologies and Quranic Tajweed (Recitation rules) ontologies. The Quranic theme-based ontologies describe the concepts and relationships that exist for the known concept in the Quran like the subject of fasting mentioned in the Quran. From the literature, it is evident that there is a significant amount of work done in developing an ontology based on different concepts mentioned in the Quran, but the literature lacks studies related to the development of the Tajweed ontology.

Theme based ontologies for certain subjects mentioned in the Quran include themes such as 'fasting' [27], 'Angels', 'the Unseen' and 'Allah (God)' in the verses of the Quran [28]. In addition, theme-based classification of concepts related to 'Iman' and 'Akhlaq' [29] have been found. Ontologies on different 'prophets' and messenger's lives, books, and the essence of their teaching they delivered to their nation were also developed [30]. Similarly, an ontology was developed to identify the concepts associated with the topic 'nature' described in the Quran [31] or subject related to the 'Human & Social Relationships' [32]. 'Solat', 'zakkat', 'sin', and 'reward' concepts were also represented in an ontology by reusing Leeds ontology [33]. Ontological model for 'place' [34] and 'time' [35] vocabulary in the Quran was also developed for helping linguistic researchers and Islamic learning.

Previous studies have almost exclusively focused on theme-based ontologies. Hence, there is a need for developing the ontology for the recitation rules for the Holy Quran. Only one study related to this context was found [36] which covered the 'Articulations points of the letters' and rules regarding 'Un vowel noon' and 'Tanween'. The scope of the study was limited so searching for the rules on the whole Quran gave limited results. Arabic letters that have multiple articulation points were not captured correctly. The ontology is not currently available publicly so find-ability and availability of the resource reduces the value of this work. Hence, this leaves room for more studies to experiment and initiate the development in this domain. We aim to remove the limitation of this study and extend the description of Arabic letters with their characteristics and Arabic letters grouping along with various Tajweed rules. Table 1 shows the comparison of our work with the previous study [36]

3. Overview of Tajweed Ontology

This section describes the design & modeling of the tajweed ontology. One of the most mature design methodology METHONTOLOGY was used to develop the ontology [5]. METHONTOLOGY methodology framework, discerns six phases. *Specification*(identification of purpose, scope, target users, implementation language, requirements, and use case scenarios)*Conceptualization* (Characteristics of ontology such as structure and components are defined in this phase) *Knowledge acquisition*(gather knowledge from different resources) *Formalization* (formal conceptual model is developed in OWL) *Integration* (the concepts already defined in other ontologies are considered to be reused instead of starting from scratch) *Implementation* (the formal model in OWL language). Figure 1 describes the conceptual model of the ontology for Arabic letters, their articulation points and their characteristics. Protégé [4] editor version 5.5.0 is used in developing the ontology using the OWL language. Protégé has a user-friendly graphical interface and allows Pellet, HermiT and other reasoners in facilitating the consistency testing. Ontology preserves all the articulation points of Arabic letters and their characteristics. This will benefit the learners and instructors of the Tajweed domain by improving the semantic searches and developing an automated intelligent system. The acquisition of knowledge necessary for building Tajweed ontology has been gathered from the unstructured resource i.e, a Tajweed book by "Kareema Carol Czerepinski" [3]. While there are different recitation styles accepted in the scholarly tradition of Tajweed sciences, we limit the design of our ontology, in particular the tajweed rules, based on the tradition of 'Hafs An Asim'. For annotating the basic meta data (e.g. title, description) the Dublin Core vocabulary[4] is used.

3.1. Use Cases

The user scenarios are identified as follows:

1. As a Tajweed teacher, I would try to search for a PDF resource that is describing the rules for the pronunciation of the Arabic letters. For example, I would search via browser for a detailed explanation of the Articulation points of Arabic letter.
2. As a Tajweed student, I would like to access to a video that is explaining the rules or characteristics of the Arabic letters with some examples for better understanding.
3. As a Tajweed seeker, I would search for a particular rule in Quran and see how many times that rule has been applied.
4. As a Tajweed learner, I would search in Quranic text and see which rules is applied within a particular word.

3.2. Competency Questions

By analyzing the scope and use cases of the Tajweed ontology, we developed a set of requirements mainly in the form of Competency Questions (CQ). These questions are categorized into patterns as mentioned by Ren et al. [37]. These CQs decide the scope and objective of the Tajweed an ontology. Table 2 shows the list of some of the CQs.

3.3. Major Classes

Concepts, individuals and relations are respectively modelled as classes, instances and properties (data type and object type properties). Axioms were represented using various OWL constraints. We used an ontology design patterns (Odp) that are the modeling solutions for some well-known an ontology modeling issues [38]. In particular, we exploit alignment Odp *Class Equivalence*, pre-

[3]https://islamhouse.com/en/books/396784/

[4]https://dublincore.org/

Table 1
Comparison of Our Work with the Previous Study

Key Features	**Previous Study**	**Our Study**
Modeling of Arabic Letters	Yes, but Multiple articulation points were not captured correctly	Yes, Correctly captured the multiple articulation points of the Arabic letter. We implemented the relationships between articulation points and the area of articulation. This was necessary in order to improve the modeling of Arabic letters.
Modeling of Characteristics of Arabic Letters	No	Yes
Modeling of Tajweed Rules	Modeled only the rules of "Noonsakinah and Tanween" (Modeled rules using class-restrictions property)	Modeled the rules of "Noonsakinah and Tanween", "Ghunnah", "Meem Sakinah" and "Qalqalah"
Data Linking/ODP	No	Yes
SWRL Rules	No	Yes
Annotations of Tajweed Rules	No	Yes
Evaluation	Logical Reasoner used	Hybrid Evaluation: Criteria based, Expert-based and Data-driven

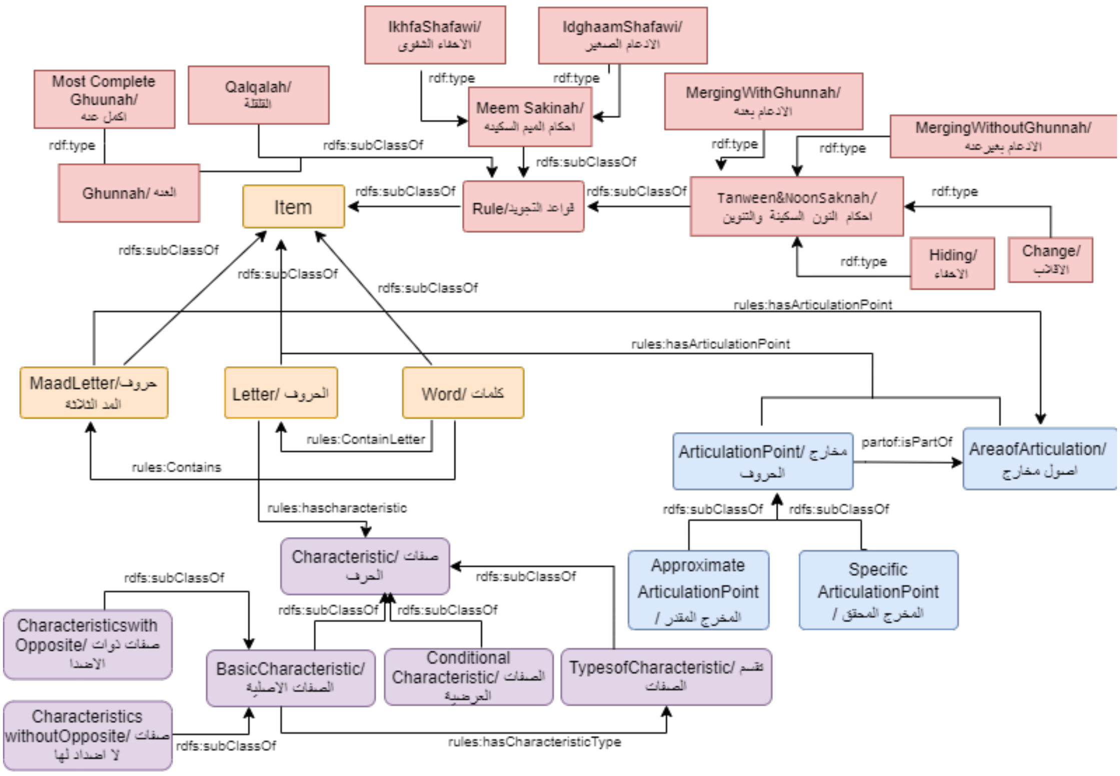


Fig. 1. T-box of Articulation Point of Letters and Letters Characteristics

Table 2

Competency Questions Mapped to Patterns with Example - CQArchetypes (CE=classexpression, OPE=objectpropertyexpression, DP=datatypeproperty, I=individual, NM = numeric modifier, PE = property expression, QM = quantity modifier)

Questions	Examples
Which [CE] [OPE] [I]?	*Which letter involves articulation point "closing two lips"?*
What is the [DP] for a particular [CE]?	*What is the linguistic definition for a particular characteristic?*
Does every [CE1] [CE2]?	*Does every letter have a characteristic?*
How many [CE1] are articulated from [CE2]?	*How many letters are articulated from the area of articulation?*
What is the [CE1] of a given [CE2]?	*What is the articulation point of a given letter?*
How many [CE] for [I]?	*How many articulation points for tongue?*
How many [CE] are there in [PE]?	*How many articulation points are there in Tajweed literature?*
What is [CE] of [I]?	*What is the characteristic type of "TheWhisper"?*
Does every [CE1] have [QM] [CE2]?	*Does every letter have min 4 characteristics?*
Mention all the [CE1] and [CE2] for a given [CE3]?	*Mention all the characteristics and types of characteristics of a given letter?*
Which are the [NM] [CE1] [CE2]?	*Which are the strong characteristic letters?*
What are the types of [CE]?	*What are the types of articulation points?*

sentation patterns *Naming patterns and Annotation patterns* and ContentODP *PartOf*. We describe four main classes that is *rules:AreaOfArticulation*, *rules:ArticulationPoint*; It has been further divided into *rules:ApproximateArticulationPoint* and *rules:SpecificArticulationPoint*. Each letter has an articulation point and each articulation point is part of some area of Articulation. The *rules:Characteristic* has two subclass *rules:BasicCharacteristic* and *rules:ConditionalCharacteristic*. *rules:Item* contains the subclasses of *rules:Letter*, *rules:MaddLetter*, and *rules:Words*. The other two classes *rules:GroupOfLetters* and *rules:TypesOfCharacteristic* were used to infer, which letter belongs to a particular group and what type of characteristics(either week, strong or middle) they have. Total 27 classes and 109 instances were developed. Classes, instances, and data-type properties were labeled in Arabic, English, and Urdu. Figure 2 shows the classes and properties used for describing the Articulation points and characteristics of Arabic letters along with the Tajweed rules.

3.4. Modeling Decisions

In the ontology, classes, instances and the relationships between them were logically modelled with an emphasis such that modeling decisions do not violate the Tajweed domain. We model the 17 articulation points (مخارج الحروف) and the 4 areas of articulation from where the 29 Arabic letters are pronounced, according to the Tajweed literature.

3.4.1. Articulation Point

Modeling the concept and relationship between an Arabic letter, its articulation points and area of articulation, we made a critical decision on how to capture those letters that have more than one articulation point. One approach was to make a compound articulation point that contains multiple articulation points. For example, one approach was to model *"rules:GumLineofFrontTwoTeeth touches the TopofTipofTongue"* as one instance. Another way to model was to use "hasArticulationPoint" object property twice with each articulation point. Example: the letter "ط" *hasArticulationPoint* "rules:GumLineofFrontTwoTeeth" and *hasArticulationPoint* "rules:TopofTipofTongue" . But these approaches have a drawback that the distinct point of Tajweed rules which are reported in the Tajweed literature can not be captured in this way. We therefore, modeled the composite articulation point in the following manner: we created another property *involvesArticulationPoint* that Joins the two articulation points of a letter with a new point (i.e. Articulationpointx). These are illustrated in Figure 3.

Another important decision was regarding the modelling of an articulation point *"rules:Teeth"*, which is used in the pronunciation of some letters. Tajweed literature goes down to the granularity of detailing the 32 teeth and their role in the pronunciation of difference Arabic letters. However, we adopt a medium granularity for modelling the different teeth and their role in the articulation points. We model Teeth as one Articulation Area, and model different areas within using the partOf relation. Moreover, for the articulation point of *"rules:throat"* we classify the parts of the throat to distinguish from which particular part of the throat the letter is articulated. Figure 4 shows the

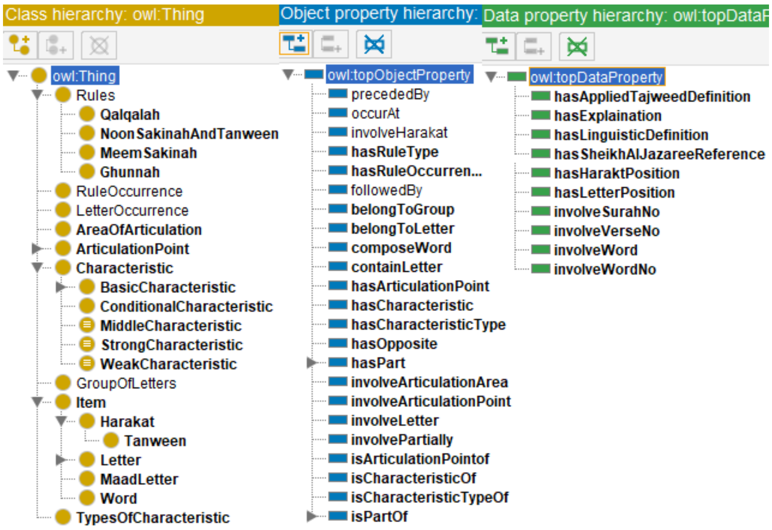


Fig. 2. Class Hierarchy, Data-type Property and Object-type Property Hierarchy

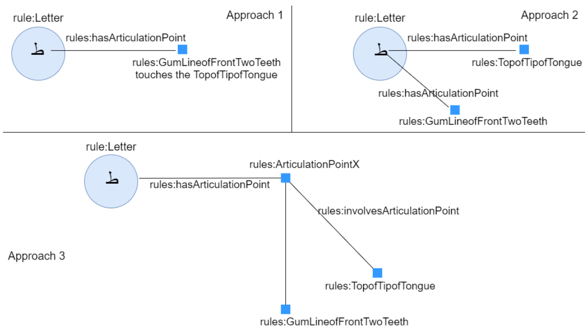


Fig. 3. Modeling Decision for Handling Composite Articulation Points

modeling of the articulation points of the teeth and the throat.

Logically modeling the definitions of classes *"ArticulationPoint"* and *"AreaOfArticulation"* as: Letter ∩ AreaofArticulation = ∅ and Letter ∩ ArticulationPoint = ∅ so, Letter *hasArticulationPoint* (ArticulationPoint ∪ AreaOfArticulation). Some areas of articulation point can be treated as an articulation point. Figure 5 shows the letter ن needs articulation point *"rules:TopofTipofTongue"* and *"rules:OppositetotheGumofTwoTopFrontIncisors"* along with an area of articulation *"rules:Nose"* for completing its pronunciation.

3.4.2. Characteristics

There is a lot of emphasis in the Tajweed literature and teaching method upon the characteristics (صفات) of the Arabic letter. Some letters in Arabic have the same articulation point, however, they differ in their characteristics. A reciter must therefore master not only the correction articulation point for the origin of the sound, the correct characteristic must also accompany the articulation. For representing the characteristics of the letter, the class *rules:BasicCharacteristic* have eight characteristics in pair (صفات متضاده). Each characteristic of the pair is opposite to the other characters in pairs. There are also seven singular characteristics that have no opposite known as *rules:CharacteristicsWithoutOpposite*(صفات غيرمتضاده). Each letter has a minimum of four characteristics, one of each pair of the opposite. We classify the characteristics into different types *rules:TypesOfCharacteristic*: The strong, the weak and the middle characteristics. Defined classes were created for weak, middle and strong characteristics and all the characteristics were inferred in their respective type. Similarly, to classify the letters according to their group type defined classes were made. Figure 6 shows the characteristics of letter " م " (rules:TheApparent, rules:TheLowered, rules:TheInBetween, rules:TheNasalization) and the type of those characteristics (rules:TheInBetween *rules:hasCharacteristicType* rules:TheMiddleCharacteristics). Moreover, it also shows the letter "م" belongs to rules:HaroofAshshafawiya.

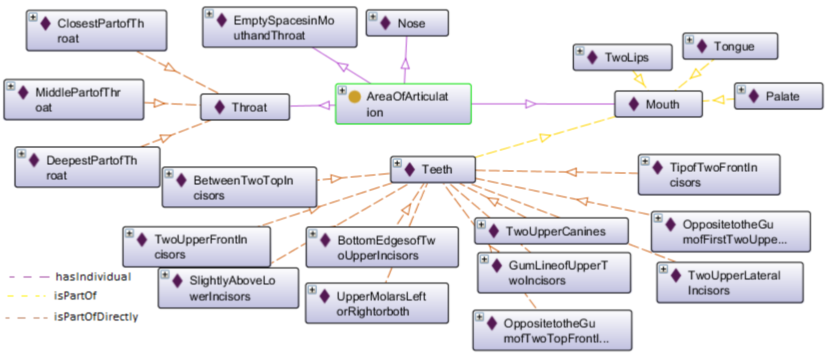


Fig. 4. Modeling of the Articulation Points Teeth and Throat

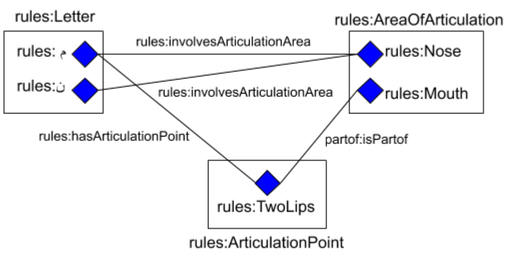


Fig. 5. Modeling Decision for Defining Articulation Point and Area of Articulation

3.4.3. Letter and Rule Occurrences

Tajweed is specific to the Quranic recitation and the rules only occur on the Quranic script. Therefore, to infer the rules on the Quranic text we first modeled the letter occurrences and rule occurrences. A letter occurence in the context of our ontology is meant to refer to a specific occurence of any given letter, within the Quranic script, with a specific location, given by the verse and the chapter(Surah) location it appears in the Quran. We separately model a letter and a diacritic. Diacritics are the marks that are placed below or above over the letter. For example a letter (ب) has a diacritic (ِ) so for modeling the letter occurrences we had two choices to represent. One approach was to take the letter with a diacritic as one instance (بِ) and model letter occurrence بِ as: *"rules:involveLetter"* (ب)nd *"rules:involveDiacritic"* (ِ) as shown in Figure 7 (option a). The second option was to model the letter and the diacritic separately. Both approaches were correct but we preferred the second approach to avoid redundancies as by adopting the first approach we have to create another class of letters with harakat (diacritic).

After modelling the letter occurrences we could easily infer the rule occurrences on each letter. Rule occurrences contain information about the rule type and the letter on which it occurs. These inferences are explained in section 3.6.

3.5. Relations in the Tajweed Ontology

This section provides a description of some of the commonly used relationships in the Tajweed ontology: is-a, partOf, hasOpposite, involvesArticulationPoint. Some classes have an "is-a" relation with other classes. For example, rules:BasicCharactersistics *is a* rules:Characteristics. The ontology employs a number of other relations including "part of". Figure8 shows the relationship between articulation point and area of articulation. Each rules:ArticulationPoint *ispartOf* rules:AreaOfArticulation. The area rule:Mouth is further divided into rule:Tongue, rule:Teeth, rule:Palate and rule:TwoLips with "partOf" relation. Relation "hasOpposite" is used for describing the instances of Characteristics. For example, in the Tajweed literature, the characteristic rules:TheApparent (الجهر) [5] *hasOpposite* rules:TheWhisper(الهمس), which means that there are some letters with the characteristic "Apparent" and the rest of the letters will have a characteristic "Whisper".

We used *owl:sameAs* property to indicate that the two URI references actually refer to the same

[5]It is a characteristic that shows there is no flow of breath as the letter is pronounced.

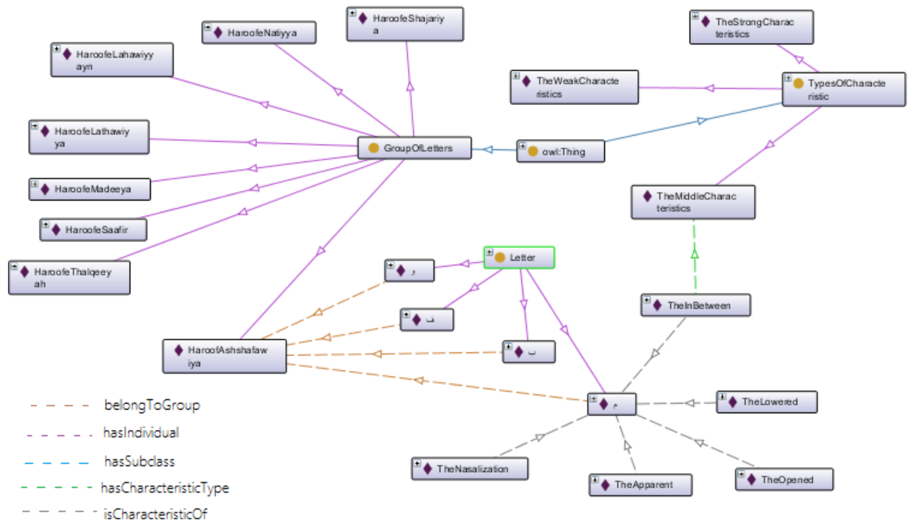
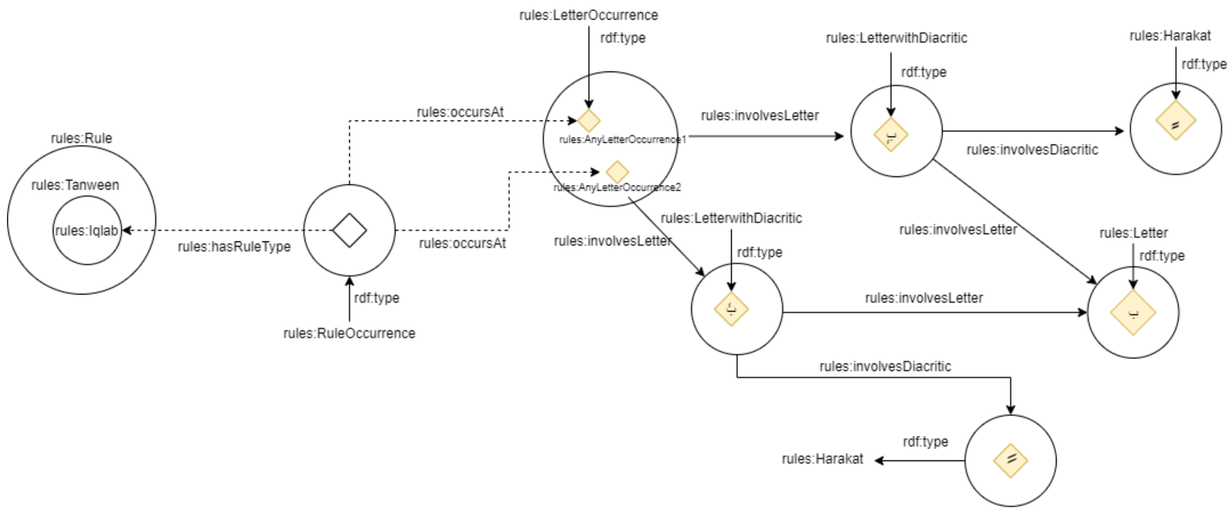


Fig. 6. Type of Characteristics and Group of Characteristics

a- Shows the modeling of letter occurrence as letter with diacritic as one instance.



concepts. For example, The concept "Articulation-Point" is the same as mentioned in the dbpedia.

```
<owl:Class rdf:about="http://www.ontology.org/
ontologies/rules#ArticulationPoint">
<owl:sameAs rdf:datatype="http://www.w3.org/
2001/XMLSchema#anyURI">
http://dbpedia.org/page/Place_of_articulation
</owl:sameAs>
</owl:Class>
```

After identifying the major Tajweed classes, and by explaining the semantic relationships between them, we also identified the domain, range, data type property, and inverse properties. Table 4 and 3 shows the list of all the relations used in the ontology.

3.6. SWRL Rules

After modeling the letter occurrences we implemented the Tajweed rules of *"NoonSakinah and Tanween", "Meem Sakinah", "Qalqalah" and "Ghunnah"* in a Semantic Web Rule Language (SWRL). It is the rule language of the semantic web, which is used to reason about OWL individuals and to infer new knowledge about those indib- Shows the modeling of letter occurrence representing letter and the diacritic separately

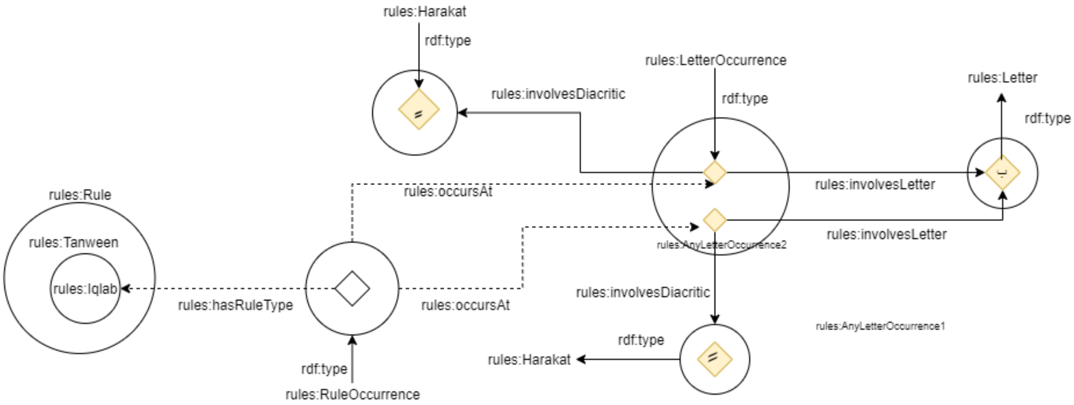


Fig. 7. Modeling Decision of Letter Occurrences for Tajweed Rules.

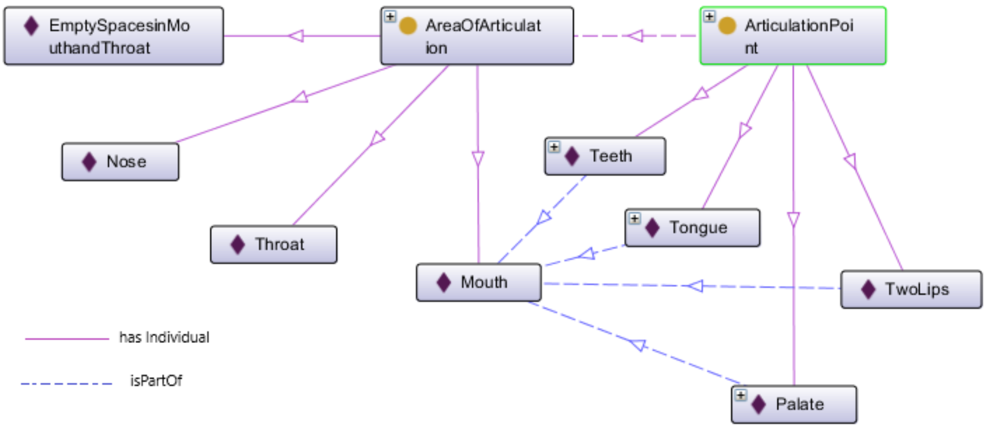


Fig. 8. Relationship between Articulation Point and Area of Articulation

viduals. SWRL includes a high-level abstract syntax for Horn-like rules, and follows this syntax: antecedent → consequent. This form means that the consequent must be true when the antecedent is satisfied. In the SWRL rules, ?x is a variable, → means implication and for conjunction ∧ symbol is used. A symbol without the leading '?' denotes the name of an instance (an individual) in the ontology. These rules provide additional expressiveness to OWL-based ontologies. The prefix "swrlx" in the rules shows the SWRL extensions built-in library. It allows users to directly create new individuals in a rule. In our rules, we build instances of rule occurrences at the run time using "makeOWLThing". This built-in provides a controlled way of creating OWL individuals in a rule. It takes two or more arguments. The first argument should be an unbound variable; an OWL individual created by the built-in will be bound to this variable. The second and subsequent arguments represent a pattern. For example, the built-in atom swrlx:makeOWLThing(?R, ?LO) will make an individual to be created and bound to ?R (Rule Occurrence) for every value of variable ?LO (Letter Occurrence). Table 5 shows the rules of Qalqalah and Iqlab rule in SWRL language with their explanation.

4. Annotated Quranic Dataset Using Tajweed Ontology

In this phase, the formal model was implemented in an ontology language. The standard ontology languages that are understandable by computers such as Web Ontology Language (OWL)

Table 3
Object-type Relations in Tajweed ontology

Object Property	Domain/Range	Characteristic	Inverse Of
rules:belongToGroup	domain:Letter range:GroupofLetters		rules:belongToLetter
rules:containLetter	domain:Word range:Letter	Asymmetric Irreflexive	rules:composeWord
rules:hasArticulationPoint	range:Letter		rules:isArticulationPointof
rules:hasCharacteristic	doamin:Letter range:BasicCharacteristic		rules:isCharacteristicOf
rules:hasCharacteristicType	domain:TypesofCharacteristic range:BasicCharacteristic		rules:isCharacteristicTypeOf
rules:hasOpposite	range:CharacteristicswithOpposite	Symmetric	
partof:hasPart	domain:AreaofArticulation range:ArticulationPoint		partof:isPartOf
rules:hasPartDirectly	domain:ArticulationPoint		rules:isPartOfDirectly
rules:involveArticulationArea	range:AreaofArticulation		rules:involveLetter
rules:involveArticulationArea	range:AreaofArticulation		
rules:involveArticulationPoint	range:ArticulationPoint		
rules:involvePartially	domain:CharacteristicswithOpposite		
rules:occurAt	domain:RuleOccurrence range:LetterOccurrence		
rules:precededBy	domain:LetterOccurrence range:LetterOccurrence		
rules:followedBy	domain:LetterOccurrence range:LetterOccurrence		
rules:hasRuleType	domain:RuleOccurrence range:TajweedRule		rules:hasRuleOccurrence

Table 4
Data-type Relations in Tajweed ontology

Data Property	Range	Annotation
appliedTajweedDefinition	range:xsd:string	rdfs:label rdfs:comment
explaination	range:xsd:string	rdfs:label
involveSurahNo	range:xsd:integer	
involveWord	range:xsd:string	
involveVerseNo	range:xsd:integer	
linguisticDefinition	range:xsd:string	rdfs:label rdfs:comment
sheikhAl-JazareeReferences	range:xsd:string	rdfs:label rdfs:comment

[39], RDF(S) [40] were used so that it could be used by semantic applications. The T-box and A-box were created for the Tajweed rules and Arabic letters and their characteristics were developed in protege editor. The rules were defined using Semantic Web Rule Language (SWRL) [41]. We use OWL API [42] as it is Java based high level Application Programming Interface (API) that supports the creation and manipulation of OWL Ontologies; and the use of reasoning engines. To modify the generic domain ontology and test the rules on the Quranic text we generated Java code of our ontology and integrate the OWL API libraries in our eclipse project. Figure 9 shows the procedure

Table 5
SWRL Rules for Iqlab And Qalqalah

Rule Name	SWRL Rule	Explanation
Iqlab	LetterOccurrence(?LO) ∧ involveLetter(?LO, ن) ∧ involveHarakat(?LO, ْ) ∧followedBy(?LO, ?LOF) ∧ LetterOccurrence(?LOF) ∧ involveLetter(?LOF, ?L) $^{I}qlabLetter(?L) \wedge hasLetterPosition(?LO,?P)$ ∧ involveSurahNo(?LO, ?S) $^{i}nvolveVerseNo(?LO,?V)$ ∧ swrlx:makeOWLThing(?R, ?LO, ?LOF) ->RuleOccurrence(?R) ∧ occurAt(?R, ?LO) ∧ hasRuleType(?R, Iqlab) ∧ hasLetterPosition(?R, ?P) ∧ involveSurahNo(?R, ?S) ∧ involveVerseNo(?R, ?V)	This rule implies that if a letter occurrence (LO) involves a letter ن and have a diacritic 'sakinah' and this letter occurrence has some verse number, surah number and has some position also LO is followed by any other letter occurrence (LOF) and LOF involves a letter ب (which is an iqlab letter) then a rule of iqlab will occur at that particular letter occurrence (LO) and shows the surah number, position of the rule and verse number.
Qalqalah	LetterOccurrence(?LO) ∧ involveLetter(?LO, ?L) ∧ QalqalahLetter(?L) ∧ involveHarakat(?LO, ْ) ∧ involveSurahNo(?LO, ?S) ∧ involveVerseNo(?LO, ?V) ∧ hasLetterPosition(?LO, ?P) ∧ swrlx:makeOWLThing(?R, ?LO) ->RuleOccurrence(?R)' ∧ occurAt(?R, ?LO) ∧ hasRuleType(?R, Qalqalah) ∧ hasLetterPosition(?R, ?P) ∧ involveSurahNo(?R, ?S) ∧ involveVerseNo(?R, ?V)	This rule implies that if a letter occurrence (LO) involves any Arabic letter (L) and it has a diacritic 'sakinah' and has some verse number, surah number and a position of each LO then a rule of qalqalah will appear at that particular letter occurrence (LO) and shows the surah number, position of the rule and verse number.

for making the annotation of the text of the Quran using Tajweed ontology.

4.1. Architecture Description

The Quran text was read verse by verse and the instances were written into the Tajweed ontology through the tajweed factory. Syntactical errors which were related to the script of the Quran were handled at the reader level. We created the automated ontology populator for the annotation purpose. All the instances were inserted into the Tajweed ontology by using tajweed factory. Tajweed factory contains all the instances, classes and relationships of our ontology. Rule engine inferred the rules present in the rule base and added them in the output OWL file where the rules occurrences were written along with the letter occurrences details. While reading the Quranic script, data preprocessing was performed to ensure the all the inconsistencies are removed from the data before the Tajweed ontology is set for validation.

Data in the real world is noisy (may contain outliers) or is incomplete means missing or unknown values. So we clean the data by identifying the patterns, correct the inconsistent data, and fill in the missing values. The outlier in our case was the absence of the diacritic "sakinah". When validating the results on the Quranic text, the rule of "Noon-Sakinah" and "MeemSakinah" was not predicted by the knowledge model because of the absence of the diacritic on Arabic letter ن and م. So, we added a sakinah on these letter. Another pattern that we identified was the presence of a Harakat " ۡ " (small upper meem) on a letter "ن" which we replace it with a diacritic "sakinah" to identify the rule of 'Iqlab' on a letter "ن" followed by the letter "ب". Also the silent letters were treated while cleaning the data. All these changes do not effect the linguistic or phonetic characteristic of the word.

After the data cleaning, the instances of the verses of the Quran were added into the ontology and annotated Quranic dataset was created. For example, the verse in the Quran [مِن شَرِّ ٱلْوَسْوَاسِ ٱلْخَنَّاسِ] contains the word مِنwithout a diacritic sakinah. In the data cleaning process, a diacritic was added on this word and all the other words along with this new word مِنْ were added into the ontology making it a populated ontology. Ontology population is a process of inserting concept and relation instances into an existing ontology which will enhance the performance and the accuracy of the knowledge model. We make changes in the A-box of our ontology by removing and adding the instances of the Quran text but the T-box remains unchanged making the vocabulary of the domain intact. The rule base contains the Tajweed rules modeled and defined in SWRL language. We run the rule engine to fire the SWRL rules that were present in the rule base to infer the rule occurrences. These rule occurrences contain the *"Letter Occurrence", "Rule Type", "Letter Position", "Verse Number" and "Surah Number"*. Thus, through this process, the whole text of the Quran was annotated. This annotated version of the Quran was in OWL format. F1 score was used to find the accuracy of the annotated ontology model. We also validate the annotated model

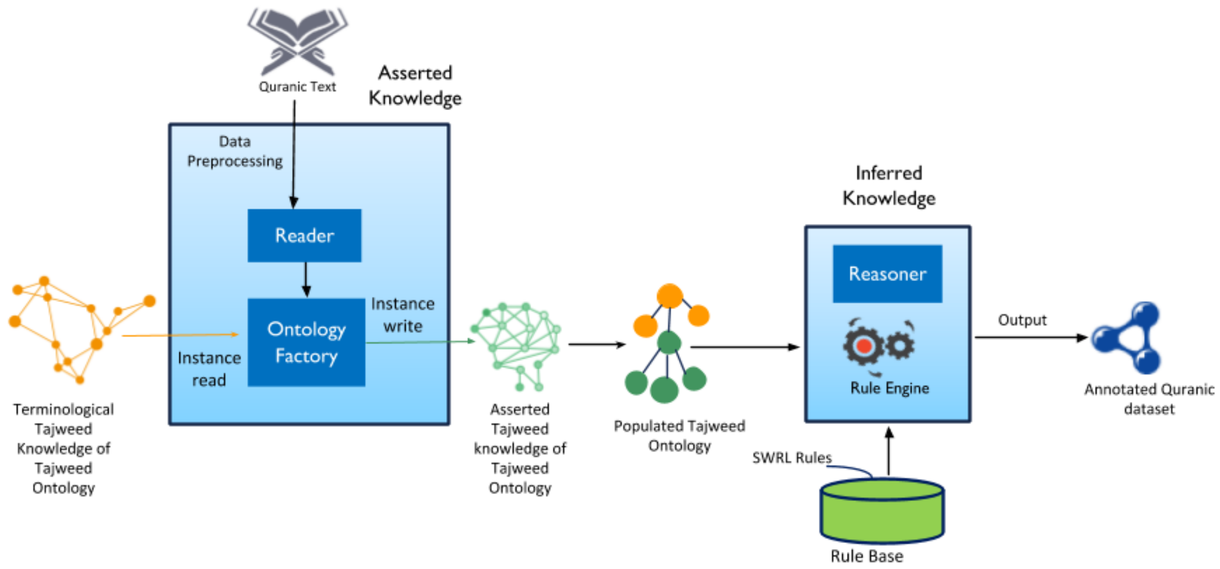


Fig. 9. Architecture for Annotating the Quranic Text using Tajweed Ontology

using the competency questions that we created in the specification phase.

4.2. Annotation and Rule Inference Example

In this section, we have described a running example of how the inferences are being deduced on the Quranic text. Figure 10 shows the knowledge is inferred on the word اِقْرَأْ (iqra). The *rules:RuleOccurrence* instances are saved into OWL format which contains the information of *rules:SurahNo*, *rules:VerseNo*, *rules:LetterPosition*, *rules:LetterOccurrence* and the *rules:Rule*. On the word اِقْرَأْ the rule of Qalalah appeared and the Letter occurrence contains the information regarding the letter and the harakat. It also contains the information regarding the preceding and followed by letter.

5. Evaluation

We employed a hybrid approach for the ontology evaluation: criteria-based [43] approach was used for evaluating the structure of the ontology, and for the correctness & accuracy human evaluation approach [44] i.e. domain experts were engaged .

5.1. Criteria-based Validation

Criteria-based approach proposed by [43] suggests a set of criteria for qualitative evaluation for the ontology. In our study, we have chosen five criteria for evaluation: Correctness, completeness, conciseness, clarity, and consistency. OOPS! (OntOlogy Pitfall Scanner!) is a method for identifying the most frequent ontological pitfalls [45]. It provides an indicator: minor, important, critical for each pitfall. It is a web-based method for assessing the ontology structure. Forty pitfalls are discussed on the basis of various criteria in this tool. These pitfalls are grouped into different groups according to different criteria. The pitfalls in the ontology are elaborated in the Table 4.

Consistency: Logical consistency was evaluated by executing the Pellet and HermiT 1.4.3.456 reasoner. It means that the concepts should be logically consistent and prevent any confusion or contradiction. The following pitfalls were checked: P05, P06, P07, P19 and P24 using OOPS!. There were no consistency issues in our ontology.

Conciseness: According to Gómez-Pérez [46] an ontology is concise a) if it does not store any unnecessary or useless definitions, b) if explicit redundancies do not exist between definitions and c) redundancies cannot be inferred using other definitions and axiom. For this evaluation criteria, the following pitfalls were checked: P02, P03, and P21.

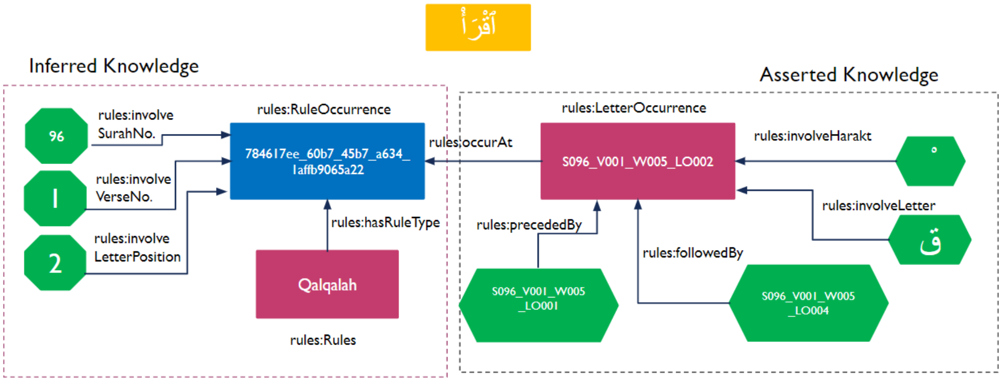


Fig. 10. Inference of Tajweed Rule on Word اَقْرَأْ

Table 6

Shows the pitfalls in Tajweed Ontology

Criteria	Pitfalls
Completeness	P11: Missing domain or range in properties
	P13: Inverse relationships not explicitly stated.
Conciseness	Your ontology does not contain any bad practice detectable by OOPS!
Consistency	Your ontology does not contain any bad practice detectable by OOPS!
Clarity	P08: Missing Annotations
	P22: Using different naming conventions
Correctness	P04: Creating unconnected ontology elements

Our ontology does not have any issues of redundancy and was marked as a concise ontology.

Clarity: Gruber [47] and Vrandečić [48] defines clarity as to whether an ontology effectively communicates the intended meaning of its defined terms and contains objective definitions that are independent of a particular context. Pitfall P08 and P22 are evaluated for this purpose. There are some wrong naming conventions added in the ontology, as we were using the camel notation, "hasPart_directly" object type property was changed into "hasPartDirectly". Some annotations were also missing so we revised our ontology by adding those annotations.

Completeness: P04, P10, P11, P12, and P13 pitfalls are used to check the completeness of the ontology. Some of the inverse relations were not mentioned in the ontology like the inverse of "belongToGroup" was missing. We created "belongToLetter" as its inverse. We revised the Tajweed ontology on the basis of the errors shown by the OOPS!. The missing domain and ranges were also revised. Gruninger et al. [49] proposed a method to check the completeness of an ontology with respect to a set of competency questions. As CQs include the requirements to be addressed by ontology we evaluated the completeness of ontology by using SPARQL queries [40].

Correctness: According to Vrandečić [48] and Gómez-Pérez [46] this criteria determines the asserted knowledge in the ontology agrees with the expert's knowledge about the domain. A higher accuracy will typically result from correct definitions and descriptions of classes, properties, and individuals. Pitfall P04 and P10 are used to check this criteria. The pitfall P04 "creating unconnected ontology elements" occurred because the class "item" was created isolated, with no relation to the rest of the ontology because this class is the collection of "Letter", "Word" and "MaadLetter". This class is not a term in Tajweed literature rather it contains terms that are used in the Tajweed domain.

5.2. Expert Evaluation

In this approach, we evaluated the accuracy or correctness of the developed Tajweed ontology by using the query retrieval method [40]. The accuracy was evaluated using the following formula;

$$Accuracy = \frac{Totalno.ofcorrectanswers}{Totalno.ofquestions} \quad (1)$$

The evaluation was conducted by five Tajweed experts. They were briefed about the different languages they can choose to see the answer to the query. The majority of the experts were selected based on Tajweed certifications obtained from a well established institute. All experts were comfortable with Arabic, English and the Urdu language. Before starting the testing and evaluation phase, a demo was shown about the whole evaluation process. The evaluation phase was recorded with the permission of the domain expert and testing sessions last for 35-40 minutes. We evaluated the ontology based on the questions expressed in natural language by experts and the competency question created in the specification phase.

A total of 16 competency questions (CQs) along with 78 questions asked by the five experts were expressed in natural language were evaluated. The ontology content was corrected using the expert's answers. Experts brief the answers to the set of questions (in natural language), the ontology should respond correctly. These questions were translated into SPARQL. Out of 94 questions 9 were incorrect. The incorrect results of the queries were revised before evaluating them with next experts. The majority of the errors were regarding the wrong annotations of the articulation point in different languages that were corrected incrementally. Expert 1 asked 11 questions other than the competency questions out of which five (5) were incorrect. Three (3) of them were regarding the wrong annotations of the articulation point in the Urdu language. The expert asked to display the articulation points of the letter ب ، ل ، و in all the three languages. The articulation points annotated in the Urdu language were corrected by the experts. Letter و articulation point was corrected to "circle of two lips" Expert 1, also corrected the characteristics of the letter ر . As per the suggestion of the expert, we removed the characteristics "TheSoftness" (الرخاوة) and "TheStrength"(االشدة) and added only the "TheInBetween" (التوسط) characteristics to represent the characteristics for the letter ر. The accuracy of the knowledge model achieved by evaluating with expert 1 was 81.48%.

Evaluating with expert 2 we analyzed the changes we made previously with the expert. Expert 2 validated those changes and asked 11 questions other than the competency questions out of which two (2) were incorrect. The articulation point for the letter "ا" was missing which we added later. The articulation point of letter ش was also identified as incorrect. The expert suggested changing the articulation point "Palate" to "Upper Palate". We achieved 92.58% accuracy with expert 2. Expert 3 evaluated a total of 34 questions out of which 16 were the competency questions. Two errors were encountered when evaluating with expert 3. The articulation point for letter ض as not according to the Tajweed literature so the expert suggested the changing of the articulation point "Upper Molar Left or right" to "Upper Molar left or right or both". The articulation point for letter ڤin Urdu annotation was also corrected. Expert 3 also reviewed all the changes that we already have made in our ontology by evaluating it with the previous experts. The accuracy we achieved by evaluating with the expert 3 was 94%. Expert 4 and 5 evaluated our ontology thoroughly. They checked each and every articulation point of letter and their characteristics. The accuracy of the knowledge model achieved with the expert 4 and 5 was 100%. They asked the questions related to the different types of characteristics, definitions of articulation points, and area of articulations, etc.

Table 8 shows the types of error and number of questions asked by each expert. Expert 4 and 5 thoroughly reviewed all the articulation points of letter and their characteristics which results in 100 % accuracy of the ontological model for Arabic letters and their characteristics. The focus of the questions was on the description of assertional and terminological axioms (i.e. subsumption of class expression, checking of instances, property hierarchy, etc.). Graph 11 shows the improvement of accuracy of the ontology model of Arabic letters and their characteristics when incrementally done with the 5 experts. Some of the SPARQL Queries we formulated for evaluating our ontology is shown in the Table 7.

5.3. Data-driven Validation

To evaluate the accuracy of Tajweed rules in the Knowledge model we use the following evaluation

Table 7
SPARQL Queries

Question	SPARQL Query

Given a letter, what is its articulation point of letter ت	SELECT Distinct ?ArticulationPoint ?APointInUrdu
	WHERE {
	rules: ت
	?ArticulationNode rules:involveArticulationPoint ?ArticulationPoint.
	?x owl:annotatedSource ?ArticulationNode.
	?x rdfs:label ?APointInUrdu.
	}

Given an Articulation Point display its Characteristics	SELECT distinct ?l ?Articulationpoint ?CharWithOpp ?CharWithoutOpp ?APinUrdu
	WHERE {{
	?l :hasArticulationPoint :ArticulationPointofTwoLips.
	:ArticulationPointofTwoLips :involveArticulationPoint ?Articulationpoint.
	?x owl:annotatedSource ?Articulationpoint.
	?x rdfs:label ?APinUrdu
	FILTER (lang(?APinUrdu)= ”ur”)
	}
	UNION
	{
	?l :hasArticulationPoint :ArticulationPointofTwoLips.
	?CharWithOpp rules:isCharacteristicOf ?l.
	?CharWithOpp rdf:type rules:CharacteristicsWithOpposite.
	}
	UNION
	{
	?l :hasArticulationPoint :ArticulationPointofTwoLips.
	?CharWithoutOpp rules:isCharacteristicOf ?l.
	?CharWithoutOpp rdf:type rules:CharacteristicsWithoutOpposite.
	}
	}

Display the rules, verse no and the words on which the rule occur, in a surah Al-Falak?	SELECT ?LO ?rule ?word ?verse ?surah
	WHERE {
	?ruleocc rules:occurAt ?LO.
	?ruleocc rules:hasRuleType ?rule.
	?LO rules:involveWord ?word.
	?LO rules:involveVerseNo ?verse.
	?LO rules:involveSurahNo ?surah.
	}
	ORDER BY ASC(?verse)

measures shown in the equations below:

$$P(Precision) = \frac{TP}{TP + FP} \quad (2)$$

$$R(Recall) = \frac{TP}{TP + FN} \quad (3)$$

$$F1(F - measure) = 2 * \frac{P * R}{(P + R)} \quad (4)$$

For validating the knowledge based ontology model, the Quranic text was used as a data set. It consists of 114 chapters (Surahs) and 6,236 verses (Ayahs). There are different writing styles of the Quran, we use the uthmani script as the Quranic corpus was also build using the same script. It is available on https://github.com/cpfair/quran-tajweed in JSON file. We converted the file into the CSV format for our ease and tested the knowledge based model for 114 surahs. A summary of dataset used is shown in table 9.

Table 8
Expert Evaluation

Expert	Total Questions	Errors	Total Correct Answer	Type of Error
1	27	5	22	3 mistakes were regarding Urdu annotation of letter ب ، ل ، و. 1 was regarding the wrong articulation point of letter وand another was regarding characteristics of letter ر.
2	27	2	25	1 mistake was regarding the missing articulation point of letter ا, and the articulation point of letter ش
3	34	2	32	1 error was regarding the Urdu annotations of articulation point ف and the other error was regarding the incorrect articulation point for letter ض.
4	29	0	29	No mistakes
5	41	0	41	No mistakes

Table 9
Dataset representation

Dataset Size	1.52 MB
Description	The Arabic text with small Alif and pause marks. The data is available as a JSON file with exact character indices for each rule.
Datatags	”surah”: 96, ”ayah”: 3, ”annotations”: ”rule”: ”qalqalah”, ”start”: 1, ”end”: 3
Example	ٱقْرَأْ بِٱسْمِ رَبِّكَ ٱلَّذِى خَلَقَ

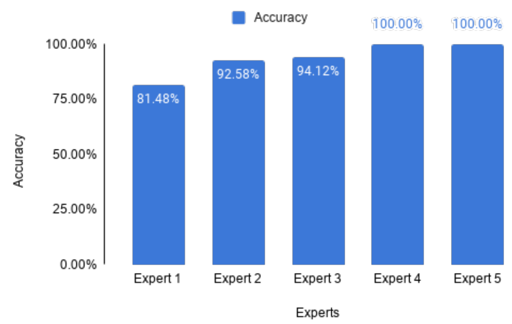


Fig. 11. Incrementally Improved Accuracy of the Tajweed Knowledge Model for Arabic Letters and their Characteristics

5.3.1. Validation for Entire Quran

We measured the F1 score for each surah (114 surahs) as it considers both precision and recall. The higher the precision (means 1), the better the system is at ensuring that the Tajweed rules are being correctly identified. Higher the Recall rate (means 1), the better the system is at not missing correct rules.

In the testing phase, we encountered some *”idgham ghunnah”* rules which were inferred by the knowledge model and were reported as False positive. We manually inspected those verses and concluded that the rule was actually correctly being inferred on occasions when the verses are treated in continuation. For example in surah 111 verse 3 [سَيَصْلَىٰ نَارًا ذَاتَ لَهَبٍ] the word ” لَهَبٍ” ends with a diacritic tanween and the next verse [وَٱمْرَأَتُهُۥ حَمَّالَةَ ٱلْحَطَبِ] starts with a letter "و". So if the reader continues to read these both verses without a pause the rule of *idgham ghunnah* will appear. In our ontology, pause marks are not dealt with.

In *surah 2-83, 85-90, 92, 97, 98 & 104* we analyse that most of the ghunnah rules were reported as false positive (FP). This was due to the fact that our ontology can not read different factors that are contributed to reading of the rules. The rule *"Most complete ghunnah"* occurs in Quran where there is a) the word contains meem shad (مّ) or noon shad (نّ) and they are coming alone means that the noon sakinah or tanween is not followed by meem shad or noon shad or the meem shad or noon shad has no tanween or sakinah on it. b) where there is idgham rule. For example, in surah[2] verse[52] [ثُمَّ عَفَوْنَا عَنكُم مِّنۢ بَعْدِ ذَٰلِكَ لَعَلَّكُمْ تَشْكُرُونَ] the most complete ghunnah rule occurs at the word "ثُمَّ" a"مِّنۢ and rule is identified at these positions because they are meem shad is coming on it own. Whereas, in surah [3] verse [158] [وَلَئِن مُّتُّمْ أَوْ قُتِلْتُمْ لَإِلَى ٱللَّهِ تُحْشَرُونَ] the word "مُّتُّمْ" has meem shadd which is followed by a noon sakinah and as meem shadd is not alone so, the rule of *"idgham with ghunnah"* should be assumed but our ontology does not cover this context so the rule of *"most complete ghunnah"* was inferred in this place.

In the Quran, there are some exceptions to words where the actual rule *"idgham ghunnah"* does not occur instead a rule "izhar e Mutlaq" exists. In our ontology, we did not cater this exception so whenever the words [ٱلدُّنْيَا صِنْوَانٌ قِنْوَانٌ بُنْيَٰنُهُۥ]come the rules of idgham ghunnah appear which results in false-negative [FN]. A future improvement would be to consider these words in the verses and obtain the perfect recall. In our ontology the FPs in *surah 1,6,9,20,21,29,57 61* happens due to the above mentioned reason.

Moreover, we also did the query validation in the second phase. We verify different types of rules and the definitions of different Tajweed concepts in the ontology model

5.3.2. Results of Data-driven Validation

For most surahs the recall measure was above 0.9 (90%) which is almost perfect recall, also the precision is overall good, as it is above 0.8 (80%) and therefore the F measure being close to 1. The lowest precision comes from the surah *48, 50, 54 and 56* in which the rule ghunnah was reported as false positive and the surah's have higher number of the mentioned rule in it. So, for the surahs with lower precision, we believe that when we account those changes which are mentioned above, the precision values of those surahs will improve even better which will result in perfect precision score. The highest precision is 1 which is achieved when all the rules in the knowledge model are correctly predicted by the data source. The perfect F score for the surahs that show less accuracy can be achieved by adding the rules that are not modeled in the knowledge base or encountered those mistakes highlighted above. The overall result of the knowledge model shows promising results as shown in Figure 12. Using equation 2, 3 and 4, the overall F1 score for the knowledge model is 0.94 (94%) and the recall is 0.99 (99%) which is nearly equal to 1. The precision value is 0.89 (89%) i.e. relevant results which were achieved by our model. These results show that the rule prediction is closer to the true/expected value. Figure 13 shows the number of the Tajweed rules predicted by the ontology based knowledge model and the data source. It shows that majority of the Tajweed rules were classified correctly in comparison with the data source. The rules of "Iqlab", "Ikhfa", "Qalqalah" ,"IkhfaShafawi" ,"IdghamwithoutGhunnah", "Iqlab", "IdghamShafawi" shows the exact result when compared them with the data source. The number of "Ghunnah" rules in knowledge model are slightly higher than the data source and the number of "idgham with ghunnah" rules are less in ontology model as compared to the data source. This difference in the number of rules results in the difference in the precision and recall value of the knowledge model.

6. Discussion & Conclusions

Our study presents the multilingual comprehensive knowledge model driven by ontology for the Tajweed of the Quran. The model comprehensively defines the vocabulary for the Tajweed domain. We modeled the articulation points for the Arabic letters and their characteristics along with

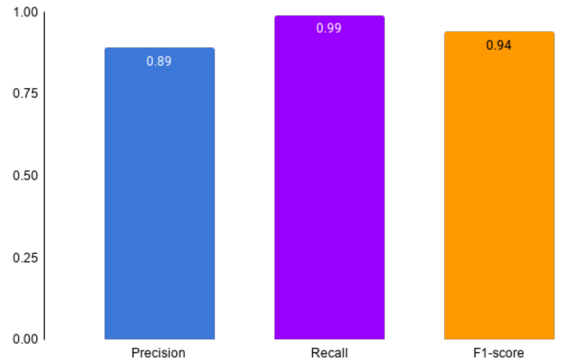


Fig. 12. Precision, Recall and F-score of the Knowledge Model

Fig. 13. Graph showing the total number of rules predicted by the knowledge base (KB) ontology model for 114 surahs VS the total number of rules in the data source

the Tajweed rules. A total of 8 rules were formalized in the model. The development of the ontology model was an iterative process. We first developed the T-box and A-box for the Arabic letters and their articulation points and validated the model by the domain experts. Our validation from the experts was also incremental. The improvement suggested by the experts were revised before evaluating with other experts. Our evaluation with the experts shows 100 % accuracy of the Tajweed ontological model. The majority of the improvements suggested by the experts were regarding the labels of the instances in different languages that should be present according to the Tajweed domain. In the second phase of development, we focused on covering the Tajweed rules. We developed the rules using the Semantic Web Rule Language (SWRL). For validating the Tajweed rules on the Quran we used precision and recall, and F-score measure. OWL API was used for parsing and rendering the OWL/XML syntax and reasoning the rule engine. The results were produced on the whole Quran (114 surahs). The overall F1 score shows 94% accuracy. By eliminating the limitations of the "Ghunnah" and "Idgham Ghunnah" rules and exception words of the Quran as highlighted in the results section, the perfect F1 score can be achieved. The results from the study show that the ontological model will provide a strong foundation to build upon and enhance the annotations of more Tajweed rules on the whole Quran.

We compared our work with the previous study [36] and observed that our work accomplished an extensive converge of Tajweed rules (total 8 rules) together with the articulation points of Arabic letters and their characteristics. We covered the multiple articulation points and the characteristics of the Arabic letters according to the Tajweed literature which the previous study fails to capture. We linked the concepts in our ontology with external sources such as "Dbpedia" and used "Dublin core" vocabulary. Moreover, our ontology follows the best practices in the ontology design area as it is multilingual and follows a sound methodology. We used SWRL Language for the development of the Tajweed rules. Furthermore, we used a hybrid approach for evaluating our ontology which not only evaluates the structure of our ontology but also evaluates the correctness & accuracy.

Our ontology creates the annotations of the Tajweed rules on the Quranic text by using the developed vocabulary, thus providing an extensive resource. We present the Tajweed rules dataset in OWL format on the entire Quran with accuracy as high as 94%. The Tajweed ontology for Arabic letters and the annotated Quran dataset using the ontology are available at https://github.com/ramshaamin/ArabicLettersOntology and https://github.com/ramshaamin/TajweedThesisV5.

While the rules covered by our ontology are not exhaustive, the representation and architecture is generic enough for more rules to be incorporated. The modelling of Letters and Rule Occurrences makes it possible for more rules to be added in the rule base.

The Tajweed ontology offers several promising use cases and potential for further research. The ontology may be utilized for annotating the vast body of resources available online, hence enhancing the semantic inter-operability and enhancing the discoverability of knowledge within these resources. In particular, educational Tajweed books, articles and multi-media resources will greatly benefit from such annotations, which include at length discussions on pronunciation of Arabic letters, their characteristics and Tajweed rules. The Quranic recitations may also be annotated using the rules annotations and these annotated datasets may be utilized for more interpretable, automated recitation assessment systems. In addition, there is also potential for extending the Tajweed ontology to model the rules in other accepted recitation styles in the Tajweed tradition.

Our ongoing work includes efforts towards incorporating more Tajweed rules (Mudood, Hamza tul wasil, Laam Sakinah) and work to eliminate the limitations of the developed version. Our architecture is generic and it will accommodate any further rules easily. We believe our study will prove to be a useful step in further development in the domain of Islamic knowledge. We also plan to expand the linked data cloud by providing a new data source of Tajweed ontology and creating richer knowledge links with available data sources. Furthermore, developing an e-Learning applications by using the ontology developed by us, will help in enhanced learning experience for the end users.

7. Acknowledgement

We are grateful for Ms. Bushra Abro and Ms. Amna Binte Kamran for their exceptional support that helps in improving our study.

References

[1] D.L. McGuinness, F. Van Harmelen et al., OWL web ontology language overview, *W3C recommendation* **10**(10) (2004), 2004.

[2] M. AMAN, A. BIN MD SAID, S.J.A. KADIR and B. BAHARUDIN, A REVIEW OF STUDIES ON ONTOLOGY DEVELOPMENT FOR ISLAMIC KNOWLEDGE DOMAIN., *Journal of Theoretical & Applied Information Technology* **95**(14) (2017).

[3] N.F. Noy, D.L. McGuinness et al., Ontology development 101: A guide to creating your first ontology, Stanford knowledge systems laboratory technical report KSL-01-05 and ..., 2001.

[4] M.A. Musen, The protégé project: a look back and a look forward, *AI matters* **1**(4) (2015), 4–12.

[5] M. Fernández-López, A. Gómez-Pérez and N. Juristo, Methontology: from ontological art towards ontological engineering (1997).

[6] D. Oberle, R. Volz, S. Staab and B. Motik, An extensible ontology software environment, in: *Handbook on ontologies*, Springer, 2004, pp. 299–319.

[7] V. Presutti, E. Daga, A. Gangemi and E. Blomqvist, eXtreme design with content ontology design patterns, in: *Proc. Workshop on Ontology Patterns*, 2009, pp. 83–97.

[8] M.C. Suárez-Figueroa, A. Gómez-Pérez and M. Fernández-López, The NeOn methodology for ontology engineering, in: *Ontology engineering in a networked world*, Springer, 2012, pp. 9–34.

[9] S. Khairuddin, S. Ahmad, A.H. Embong, N.N.W.N. Hashim and S.S. Hassan, Features Identification and Classification of Alphabet (ro) in Leaning (Al-Inhiraf) and Repetition (Al-Takrir) Characteristics, in: *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, IEEE, 2019, pp. 295–299.

[10] M. Al-Ayyoub, N.A. Damer and I. Hmeidi, Using deep learning for automatically determining correct application of basic quranic recitation rules., *Int. Arab J. Inf. Technol.* **15**(3A) (2018), 620–625.

[11] A. Qayyum, S. Latif and J. Qadir, Quran reciter identification: A deep learning approach, in: *2018 7th International Conference on Computer and Communication Engineering (ICCCE)*, IEEE, 2018, pp. 492–497.

[12] L. Marlina, C. Wardoyo, W.M. Sanjaya, D. Anggraeni, S.F. Dewi, A. Roziqin and S. Maryanti, Makhraj recognition of Hijaiyah letter for children based on Mel-Frequency Cepstrum Coefficients (MFCC) and Support Vector Machines (SVM) method, in: *2018 International Conference on Information and Communications Technology (ICOIACT)*, IEEE, 2018, pp. 935–940.

[13] A.H. Ahmed and S.M. Abdo, Verification system for Quran recitation recordings, *Int. J. Comput. Appl.* **163**(4) (2017), 6–11.

[14] A.M. Hasan, T.H. Rassem, M. Noorhuzaimi et al., Combined Support Vector Machine and Pattern Matching for Arabic Islamic Hadith Question Classification System, in: *International Conference of Reliable Information and Communication Technology*, Springer, 2018, pp. 278–290.

[15] B. Yousfi, A.M. Zeki and A. Haji, Holy Qur'an Speech Recognition System Distinguishing The Type of prolongation, *Sukkur IBA Journal of Computing and Mathematical Sciences* **2**(1) (2018), 36–43.

[16] B. Yousfi and A.M. Zeki, Holy Qur'an speech recognition system Imaalah checking rule for warsh recitation, in: *2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA)*, IEEE, 2017, pp. 258–263.

[17] A.O. Adeleke, N.A. Samsudin, A. Mustapha and N.M. Nawi, A Group-Based Feature Selection Approach to Improve Classification of Holy Quran Verses, in: *International Conference on Soft Computing and Data Mining*, Springer, 2018, pp. 282–297.

[18] S. Hazmoune, F. Bougamouza, S. Mazouzi and M. Benmohammed, A new hybrid framework based on Hidden Markov models and K-nearest neighbors

for speech recognition, *International Journal of Speech Technology* **21**(3) (2018), 689–704.

[19] M. Maqsood, H.A. Habib and T. Nawaz, An efficientmis pronunciation detection system using discriminative acoustic phonetic features for arabic consonants., *Int. Arab J. Inf. Technol.* **16**(2) (2019), 242–250.

[20] S. Bloehdorn and A. Hotho, Text classification by boosting weak learners based on terms and concepts, in: *Fourth IEEE International Conference on Data Mining (ICDM'04)*, IEEE, 2004, pp. 331–334.

[21] H. Afzal and T. Mukhtar, Semantically Enhanced Concept Search of the Holy Quran: Qur'anic English WordNet, *Arabian Journal for Science and Engineering* **44**(4) (2019), 3953–3966.

[22] R. Ahmed and E. Atwell, Developing an Ontology of Concepts in the Qur'an, *International Journal on Islamic Applications in Computer Science And Technology* **4**(4) (2016), 1–8.

[23] H. Shmeisani, S. Tartir, A. Al-Na'ssaan and M. Naji, Semantically Answering Questions from the Holy Quran, in: *International Conference on Islamic Applications in Computer Science And Technology*, 2014, pp. 1–8.

[24] A. Alfaries, M. Albahlal, M. Almazrua and A. Almazrua, A rule-based annotation system to extract Tajweed rules from Quran, in: *2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences*, IEEE, 2013, pp. 281–286.

[25] S. Silva, R. Pereira and R. Ribeiro, Machine learning in incident categorization automation, in: *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, IEEE, 2018, pp. 1–6.

[26] S. Bloehdorn and A. Hotho, Text classification by boosting weak learners based on terms and concepts, in: *Fourth IEEE International Conference on Data Mining (ICDM'04)*, IEEE, 2004, pp. 331–334.

[27] S.A.M. Zailani, N.A. Omar, A. Mustapha and M.H.A. Rahim, Fasting Ontology in Pillars of Islam, *Indonesian Journal of Electrical Engineering and Computer Science* **12**(2) (2018), 562–569.

[28] A. Ta'a, Q. Abed and M. Ahmad, Al-Quran ontology based on knowledge themes, *Journal of Fundamental and Applied Sciences* **9**(5S) (2017), 800–817.

[29] B.B.M. Ali and M. Ahmad, Al-Quran themes classification using ontology, *Icoci. Cms. Net. My* **74** (2013), 383–389.

[30] H.A. Al-Sanasleh and B.H. Hammo, Building domain ontology: Experiences in developing the prophetic ontology form Quran and hadith, in: *2017 International Conference on New Trends in Computing Sciences (ICTCS)*, IEEE, 2017, pp. 223–228.

[31] A.S. Sadi, T. Anam, M. Abdirazak, A.H. Adnan, S.Z. Khan, M.M. Rahman and G. Samara, Applying ontological modeling on quranic" nature" domain, in: *2016 7th International Conference on Information and Communication Systems (ICICS)*, IEEE, 2016, pp. 151–155.

[32] Y.M. Tashtoush, M.R. Al-Soud, R.M. AbuJazoh and M. Al-Frehat, The noble quran Arabic ontology: Domain ontological model and evaluation of human and social relations, in: *2017 8th International Conference on Information and Communication Systems (ICICS)*, IEEE, 2017, pp. 40–45.

[33] A.R. Yauri, R.A. Kadir, A. Azman and M.A. Murad, Quranic verse extraction base on concepts using OWL-DL ontology, *Research Journal of Applied Sciences, Engineering and Technology* **6**(23) (2013), 4492–4498.

[34] W. Alromima, I.F. Moawad, R. Elgohary and M. Aref, Ontology-based model for Arabic lexicons: An application of the Place Nouns in the Holy Quran, in: *2015 11th International Computer Engineering Conference (ICENCO)*, IEEE, 2015, pp. 137–143.

[35] M. Al-Yahya, H. Al-Khalifa, A. Bahanshal, I. Al-Odah and N. Al-Helwah, An ontological model for representing semantic lexicons: an application on time nouns in the holy Quran, *Arabian Journal for Science and Engineering* **35**(2) (2010), 21.

[36] E.A.A. Ibrahim, M.A.M. Ataelfadiel and E.S. Atwell, Provisions of Quran Tajweed Ontology (Articulations Points of the Letters, UN Vowel Noon and Tanween), *International Journal of Science and Research* **6**(8) (2017), 756–761.

[37] Y. Ren, A. Parvizi, C. Mellish, J.Z. Pan, K. Van Deemter and R. Stevens, Towards competency question-driven ontology authoring, in: *European Semantic Web Conference*, Springer, 2014, pp. 752–767.

[38] A. Gangemi, Ontology design patterns for semantic web content, in: *International semantic web conference*, Springer, 2005, pp. 262–276.

[39] D.L. McGuinness, F. Van Harmelen et al., OWL web ontology language overview, *W3C recommendation* **10**(10) (2004), 2004.

[40] I. Kollia, B. Glimm and I. Horrocks, SPARQL query answering over OWL ontologies, in: *Extended Semantic Web Conference*, Springer, 2011, pp. 382–396.

[41] I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, M. Dean et al., SWRL: A semantic web rule language combining OWL and RuleML, *W3C Member submission* **21**(79) (2004), 1–31.

[42] M. Horridge and S. Bechhofer, The owl api: A java api for owl ontologies, *Semantic web* **2**(1) (2011), 11–21.

[43] J. Yu, J.A. Thom and A. Tam, Ontology evaluation using wikipedia categories for browsing, in: *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, ACM, 2007, pp. 223–232.

[44] M. Grüninger and M.S. Fox, Methodology for the design and evaluation of ontologies (1995).

[45] M. Poveda-Villalón, A. Gómez-Pérez and M.C. Suárez-Figueroa, Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation, *International Journal on Semantic Web and Information Systems (IJSWIS)* **10**(2) (2014), 7–34.

[46] A. Gómez-Pérez, Evaluation of ontologies, *International Journal of intelligent systems* **16**(3) (2001), 391–409.

[47] T.R. Gruber, Toward principles for the design of ontologies used for knowledge sharing?, *International journal of human-computer studies* **43**(5–6) (1995), 907–928.

[48] D. Vrandečić, Ontology evaluation, in: *Handbook on ontologies*, Springer, 2009, pp. 293–313.

[49] M. Grüninger and M.S. Fox, Methodology for the design and evaluation of ontologies (1995).

[50] A. Basharat, K. Rasheed and I.B. Arpinar, Towards linked open islamic knowledge using human computation and crowdsourcing, in: *Proceedings of the International Conference on Islamic Applications in Computer Science And Technology*, 2015.

[51] R. Studer, V.R. Benjamins and D. Fensel, Knowledge engineering: principles and methods, *Data & knowledge engineering* **25**(1–2) (1998), 161–197.

[52] F. Neuhaus, F. Neuhaus, S. Ray and R.D. Sriram, *Toward Ontology Evaluation across the Life Cycle*, US Department of Commerce, National Institute of Standards and Technology, 2014.

[53] M. Fernández-López, Overview of methodologies for building ontologies, in: *IJCAI99 workshop on ontologies and problem-solving methods: Lessons learned and future trends*, Vol. 430, 1999.