

## Article

# A Knowledge-Based Sense Disambiguation Method to Semantically Enhanced NL Question for Restricted Domain

Ammar Arbbaeen <sup>1,\*</sup>  and Asadullah Shah <sup>2</sup> 
<sup>1</sup> Department of Computer Science, Faculty of Information and Communication Technology, International Islamic University Malaysia, Selangor 53100, Kuala Lumpur, Malaysia

<sup>2</sup> Faculty of Information and Communication Technology, International Islamic University Malaysia, Selangor 53100, Kuala Lumpur, Malaysia; asadullah@iium.edu.my

\* Correspondence: arbbaeen.ammar@live.iium.edu.my

**Abstract:** Within the space of question answering (QA) systems, the most critical module to improve overall performance is question analysis processing. Extracting the lexical semantic of a Natural Language (NL) question presents challenges at syntactic and semantic levels for most QA systems. This is due to the difference between the words posed by a user and the terms presently stored in the knowledge bases. Many studies have achieved encouraging results in lexical semantic resolution on the topic of word sense disambiguation (WSD), and several other works consider these challenges in the context of QA applications. Additionally, few scholars have examined the role of WSD in returning potential answers corresponding to particular questions. However, natural language processing (NLP) is still facing several challenges to determine the precise meaning of various ambiguities. Therefore, the motivation of this work is to propose a novel knowledge-based sense disambiguation (KSD) method for resolving the problem of lexical ambiguity associated with questions posed in QA systems. The major contribution is the proposed innovative method, which incorporates multiple knowledge sources. This includes the question's metadata (date/GPS), context knowledge, and domain ontology into a shallow NLP. The proposed KSD method is developed into a unique tool for a mobile QA application that aims to determine the intended meaning of questions expressed by pilgrims. The experimental results reveal that our method obtained comparable and better accuracy performance than the baselines in the context of the pilgrimage domain.

**Keywords:** question answering systems; knowledge-based method; natural language processing; word sense disambiguation; ontology; WordNet



**Citation:** Arbbaeen, A.; Shah, A. A Knowledge-Based Sense Disambiguation Method to Semantically Enhanced NL Question for Restricted Domain. *Information* **2021**, *12*, 452. <https://doi.org/10.3390/info12110452>

Academic Editor: Diego Reforgiato Recupero

Received: 7 October 2021

Accepted: 28 October 2021

Published: 31 October 2021

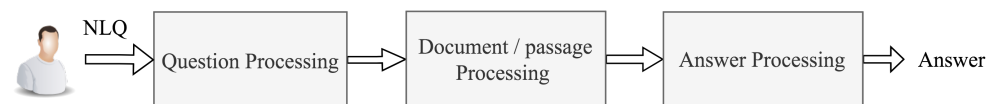
**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The emerging field of information retrieval (IR) based on question answering (QA) systems integrates the research and techniques from natural language processing (NLP), information extraction (IE), automatic summarization, knowledge representation and database systems. In QA applications, the user may obtain precise and concise answers to questions from the stored documents. While in IR applications, the user searches by keywords as input and receives a relevant list of documents based on the query [1]. The research in QA, where a system is required to understand NL questions and infer precise answers, has recently drawn considerable attention. NLP supports QA systems with understanding expressions techniques to comprehend the NL questions. On the other hand, IE utilises NLP techniques and tasks such as name entity recognition, vocabulary analysis, template elements and relationship extraction to generate structured machine-readable format from text to obtain appropriate information. Therefore, QA systems attract various researchers who are achieving significant enhancements in different domains such as biomedical, weather forecasting and tourism [2]. A QA system commonly follows the standard processing phases as shown in Figure 1.



**Figure 1.** Basic architecture of a question answering system.

The basic architecture of a QA system consists of three major modules: a question processing module, a documents/passages processing module, and an answer processing module [3]. The question processing module is the most fundamental phase in the question answering application. The quality of NL question analysis affects the performance of further modules such as answer processing. The question processing module is responsible for analysing the received NL questions, understanding and identifying the intended meaning. Analysing and classifying the NL questions are performed on the basis of different NLP approaches that range from keyword extraction to deep linguistic-based analysis [4]. The syntactic and semantic aspects of the posed questions are analysed to determine the intended meaning in an appropriate structured format. The output entities are then used by the information retrieval method to match with a relevant candidate during the processing phase. The candidates are further processed in a passage analysis module using methods such as sentence splitting, part-of-speech (POS) tagging and parsing [5]. The answer processing module is the final phase in the QA system, that uses the output analysis results to extract a precise and concise answer based on the NL question's structured representation and ranking of candidate results.

The major challenge faced by an NLP-based question processor is the usage of a natural language that is full of ambiguity [6,7]. Ambiguity is defined as a barrier to human language understanding [8]. The lexical semantic issue is mainly arising due to the difference between the KB representation of the terms and the information expressed by the input question [9]. Some of these ambiguities require deep linguistic analysis, while others are easier to determine [10]. These cause it to process more than one interpretation for an individual word. For instance, if the word “chair” is used, then it presents different lexical definitions, including ‘a seat for a person’ and ‘the officer who presides at the meeting of organisations’.

The ambiguity associated with NL questions may lead to retrieving irrelevant answers or no answers at all. A close look at the relevant literature reveals that researchers [11,12] have reported progressive results to lexical issues on the topic of word sense disambiguation (WSD). In addition, several studies [9,13] have appraised this problem in the context of question answering, however, challenging questions may require handling ambiguities based on date and location-specific to certain conditions, such as seeking instructions or advice in a critical situation [1,8,14]. For instance, a user posed a question: What are the hajj pillars? The sense of “pillars” in this context refers to “a fundamental principle or practice” defined by the WordNet 3.1. However, “pillars” in the following context “how should I reach the pillars?” refers to “anything that approximates the shape of a column or tower”, which refers to the stoning of the devil pillars at Mina especially if the question's metadata indicate the Hajj season and the location is in the area of Mina. Similarly, “what are the rulings of praying at the bank of Masjid Al-Haram while following the imam?”, the sense of the term “bank” in this context is “sloping land”. Likewise, “what are the dates available for Umrah?” the term dates in this context refer to the “day of the month or year as specified by a number”. However, the term “date” in the following question refer to the fruit type “what are the dates available from Al-Madinah?”

The ambiguity problem is widely encountered in the literature, but still requires many efforts from the research community to resolve the numerous problems involved that leads to the misinterpretation [15,16]. Therefore, the motivation of this work is to propose a new knowledge-based sense disambiguation (KSD) method that aims at resolving the problem of lexical ambiguity occur in question processing. To the best of our knowledge, none have reported the integration of question's metadata, context knowledge and ontological information into shallow NLP in resolving the problem of ambiguity especially in the

context of the pilgrimage domain. In this novel scheme, the WordNet 3.1 is used to obtain context knowledge based on surrounding words in a question. The proposed KSD method helps to determine the intended meaning and assign the appropriate sense of the polysemous words associated with NL questions.

The analysis of the literature revealed that a close domain like a pilgrimage where Pilgrims are specifically dependant on information that could be accessed on the internet. Current internet queries require exact keywords to be similar or identically structured in order to retrieve precise relevant information [17,18]. In addition, there is a lack of consideration to the role of WSD for lexical resolution in questions expressed by pilgrims to obtain critical information [19]. Therefore, there is a need to establish an effective QA approach to bridge the gap between the formal representation of the knowledge base and the pilgrim's expressiveness, which would semantically enhance the quality of queries and guarantee accurate answers provision.

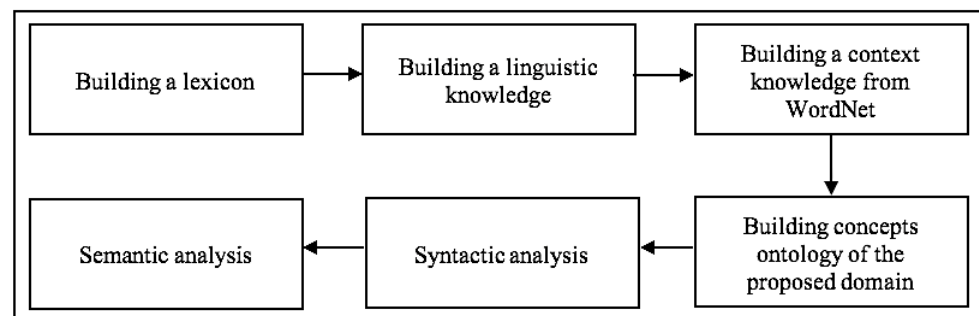
The major contributions of this study are as follows:

- This work investigates the role of question's metadata (date/GPS) in WSD to semantically enhanced question processing.
- A novel knowledge-based sense disambiguation method to solve lexical ambiguity issues in NL questions.
- A novel tool for pilgrimage question answering application based on the proposed KSD approach. This application enables pilgrims to acquire concise, accurate, and quick answers for various questions expressed in natural language anytime anywhere via their smartphones.

The remainder of this paper has been organized as follows. Section 3 introduces the related works. Section 4 presents the system architecture that includes the proposed KSD method. Section 5 concerns the evaluation of the proposed method into the QA system. Section 6 presents the discussion and consists of findings and limitations of the proposed method. Finally, the conclusions and future work are discussed in Section 7.

## 2. Methodology

The proposed KSD method is constructed in several required phases including building a lexicon and linguistic knowledge, building the context KB using the WordNet 3.1, and the design of a novel domain-specific ontology is developed using the Protégé software (V 5.5.0). These knowledge bases are integrated into the KSD to process syntactic and semantic analyses for the posed question, see Figure 2. In the present study, a syntactic analysis algorithm has been developed to chunk the question and segment its context into grammatical parts. This provides the linguistic knowledge of the question in the form of a structured format. The semantic analysis algorithm receives the syntactic structured form and process WSD for the target words by utilising the context knowledge base and ontological information. The phase also utilises the semantic role labelling (SRL) process to assign labels to the terms in the context which determine their roles and represent them in the form of a semantic structured format. A set of classification rules is used for predicting the answer type of the question, which supports the answer processing module in QA systems.



**Figure 2.** Methodology pipeline of the knowledge-based sense disambiguation method.

### 3. Related Works

A QA system aims to infer precise and concise answers in response to a user's NL question. The capabilities to comprehend NL questions and related context are crucial in QA system systems [20]. Because of the dynamic nature of a natural language, establishing a QA system has been classified as a challenging task. This is due to the ambiguity issue associated with NL questions, which affect the answer processing module [15,16,21,22]. Most of the studies that carried out in QA systems have used shallow methods like linking words in NL questions with the same words in the regained passages [2]. The requirement of actual methods could exactly return precise answers made by the community of QA to turn towards deep learning, knowledge graph, NLP techniques, etc. [23]. QA is usually considered as one of the most demanding tasks in NLP [15].

Determining ambiguity during question processing is a crucial step in QA systems. Besides, ref. [24] maintained that an NL question contains some cue expressions related to its class. They offered an approach that learns these expressions for every type of query from the social question answering collections. The template-based method for question representation over various knowledge bases and corpora was offered by [25]. It concerns mapping the posed question to the available learned templates. To perform this, the context of the NL question is substituted with its notions through the conceptualization mechanism. Utilizing such templates assists in factoid and complex questions that could be broken into a segment of questions that corresponds to one predicate. Furthermore, the semantic classification method proposed in [26] for categorising question-like search queries into a group of twenty-six pre-defined semantic classes. For the categorization, a maximum entropy model is employed, and the training set is constructed of the query sequences in the user's search sessions. In the same year, two kinds of open-ended questions, "how" and "why", were developed by [27] for the community QA system. The identification of question class is carried out by using a machine learning algorithm on the patterns of questions. Answer extraction is subsequently carried out by returning the ending and starting points of the so-called elementary units of discourse, that resemble clauses or sentences. The accuracy of answer extraction was reported 90% by manual assessment.

Identifying appropriate characteristics for answer generation or question classification is an important task. In this line, the eleven diverse characteristics extracted from the questions are semantic (called entities) features, surface and syntactic (e.g., lexical chains). The characteristics are employed in [26] to conclude the semantic class of the query. Moreover, the utilization of the genetic programming method provides features training for the QA system in [28]. A group of features based on semantic, syntactic and lexical of the sentences were defined. The sentences and paragraphs are ordered on the basis of the feature set for factoid and definitional types of questions. Another work offered a model on the basis of reinforcement training which generates a multi-document summary as an answer to a complex query [29]. They characterise the sentences in respect to static features to determine the essential sentence in a context and measure the similarity between the sentence and the question. The similarity between the candidate and the remaining sentences are also measured by dynamic features employed in the text summarization context.

In addition, the semantic representation embedding for QA aims to convert a question posed in a natural language into a logical form utilizing the Freebase database as the source of knowledge [30]. A query is represented in respect to logical features and lexical features (words), including entity type, topic, the category of the questions, and a predicate about the answers classification type. The lexical items are linked to logical characteristics and an answer is created employing the similarities of the words in the context of the question and logical characteristics of potential answers. They achieved 37% in terms of accuracy, 56% recall, and 45% F-measure on a dataset of about 2000 question-answer pairs.

In [31], a model cross attention-based neural network (NN) was tailored to knowledge base QA tasks. This takes into account the mutual impact between aspects of the corresponding responses and the representation of questions. The model influences the global KB information, which aims to precisely represent the answers. Moreover, a novel graph convolution-based neural network, named GRAFT-Net (Graphs of Relations among Facts and Text Networks), specially developed to function over varied graphs of text sentences as well as KB. To reply to a question posed in NL, GRAFT-Net takes into consideration a varied graph developed from KB, and therefore could influence the rich relational structure between the two sources of information [32].

#### *Review of the Most Related QA Systems for the Pilgrimage Domain*

Pilgrims might face some challenges during the Hajj session which need immediate decisions like getting lost, losing spatial reference or missing religious rules. A huge number of online queries are conducted to get precise information related to the pilgrimage. Nevertheless, the outcomes of such searches create hundreds of web addresses and sometimes convoluted search results. Based on [33], the existence of web platforms with a huge pool of information web pages that are produced on a daily basis has changed how users deal with the internet [1].

Several studies have been carried out to assist pilgrims in obtaining precise information, like the Islamic domain as one of the essential aims of ontology development. In their research, ref. [34] suggested a knowledge-based expert system to support pilgrims at several stages such as ritual exercises learning and decision-making. In the meantime, ref. [35] offered a semantic system on the basis of a QA which make it possible for pilgrims to generate questions about Umrah in the form of NL. The proposed system utilized concept ontology to represent the knowledge base of the ritual dedicated for the Umrah related exercises. Several difficulties of NL questions were noticed in the system. This is due to the fact that all questions had to be linked with the contents in the ontology. The Umrah educational module is among the most current existing systems. Likewise, a dynamic knowledge-based method which simplified a decision support system employed by pilgrims to enquiry information from dynamic knowledge-based records [36]. Besides, the system makes it potential for pilgrims to interact with experts, to get convenient possible resolutions to their enquiries. In the meantime, the application called M-Hajj DSS offered advanced and simple questions as well as answers utilizing artificial intelligence (AI), on the basis of decision trees and case-based reasoning (CBR). Nevertheless, the system was created for Android smartphones, using the Malay language, and also failed to take into account the ambiguity issue associated with user expression. The language issue was dealt with by [37], who provided a mobile translation application to make it possible for pilgrims to complete rituals that included various languages and sought to help pilgrims to interact effortlessly with other pilgrims who speak multiple languages.

In conclusion, the analytical research carried out by [19] revealed that the developers of the majority of recent pilgrimage applications are interested in essential usage, and their applications do not have interactive features, with users focused on visual applications, and more than 87% of these provide just one language. The research showed some defects in presenting certain interactive features which assist practical communication among users, like fatwa chatting, and revealed that current pilgrimage mobile applications do not have real-time map updates to help pilgrims, i.e., awareness of location. Nonetheless,

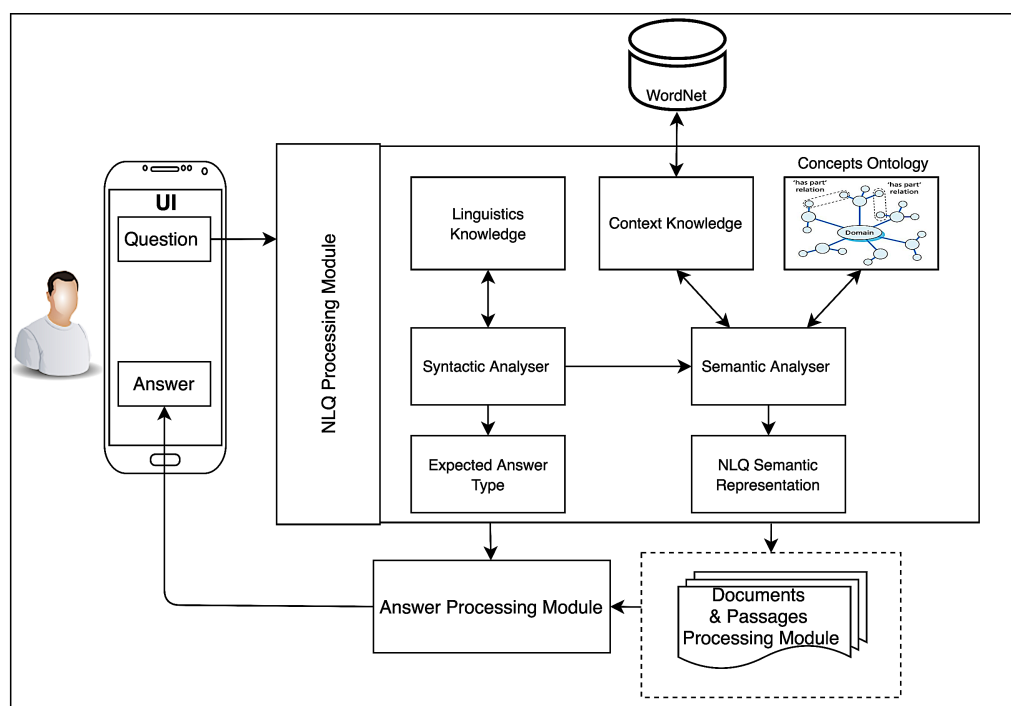


no extant study has taken into consideration the WSD role in sorting out lexical-semantic vagueness in expressive NL questions by pilgrims to increase the QA accuracy performance in the domain.

In a nutshell, the following required features are used for the sense disambiguation method to bridge the gap of lexical issues in the context of NL questions posed in a QA application. In this work, a QA application prototype based on WSD method has been developed. Due to the previously mentioned drawbacks, the knowledge-based method is the most appropriate method to be utilised in NL questions processing in the context of the pilgrimage domain, as this method incorporates multiple knowledge sources context knowledge obtained from the WordNet 3.1, ontological information, and question's metadata. For this method to work, domain knowledge was required, which was modelled on ontology. Utilizing knowledge about real-life was categorized into an AI-based method; this method was symbolised by using lexicon domain ontology. Such ontology provides rich knowledge and facts in which every concept was connected to its meaning. Moreover, the QA system with a meaningful constituent is provided by the question's metadata integrated with the ontology. "These set of rules are created to define the exercise a user is performing at a specific location and time of the year, which ultimately assist in generating some semantic structures for the context in the NL question. To use this method, a shallow-based method was established which works in semantic and syntactic levels [1].

#### 4. System Architecture

In this section, we present the fundamental architecture of the QA application based on KSD as revealed in Figure 3. The basic QA system consists of three major modules: NL question, documents, and answer modules. This study is focused only on the NL question processing and the answers modules. The NL question module integrates into the user interface (UI) of the QA application. The question posed by the user is presented at the UI of the mobile application and at which point passes through a sequence of processes.



**Figure 3.** The architecture of the QA system based on KSD.

#### 4.1. User Interface

The user interface (UI) allows the user to pose a question in natural language and obtain concise accurate information. This study focuses on the wh-form of questions in the English natural language within the context of the pilgrimage domain. The design of the UI requires the utilization of tools, dialogues, and buttons for posting NL questions and receiving answers. This study examines the role of WSD as a proposed method to identify the accurate meaning of lexical ambiguity associated with questions posed by pilgrims through the mobile application UI. KSD is applicable when the question contains predicate-argument structures. The proposed method is depicted in the NL question processing module, which is an integral constituent of NLP illustrated in Figure 4.

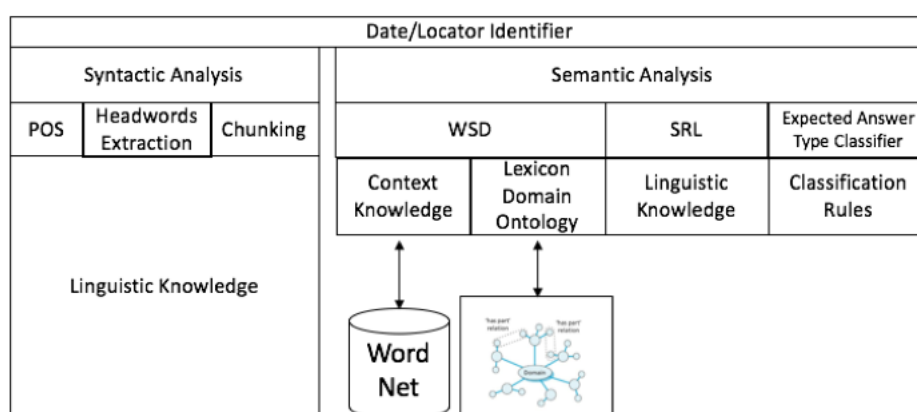


Figure 4. The proposed KSD approach.

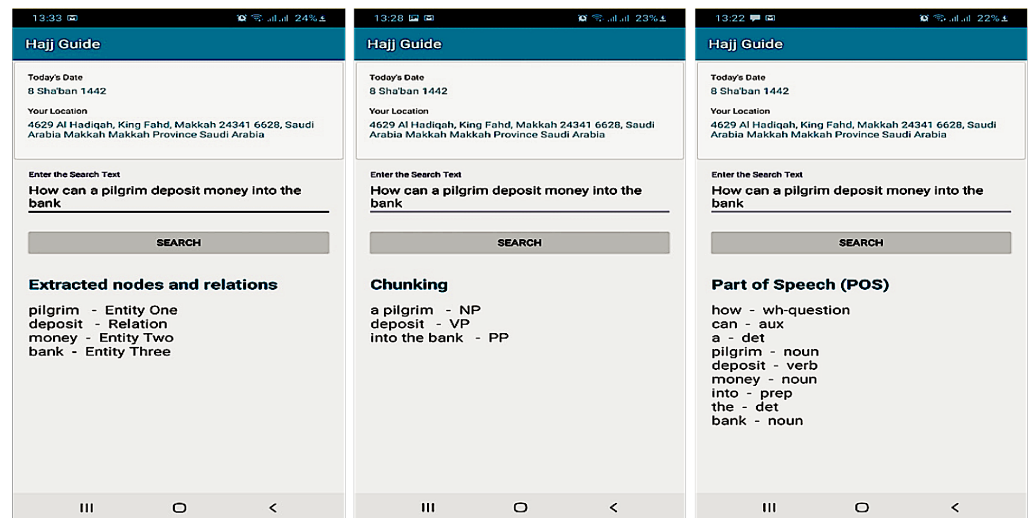
#### 4.2. Syntactic Analysis

The aim of the grammatical analyser engages the capability to analyse a question (group of words) that identify the parts of a question (preposition, adjectives, verb, noun, etc.). The link of a sentence component to a higher-order unit that has the ability to perform discrete grammatical meanings (verb groups, noun groups or phrases, etc.). The NL question processor firstly performs the task of POS in the syntactic analysis module. The suggested rule-based tagging assigns potential tags to every lexical item in the NL question. For instance, 'How can a pilgrim contact an agent in this area?', could be labelled as:

[How/Wh-Q] [can/Aux] [a/Det] [pilgrim/Noun] [contact/Verb] [an/Det] [agent/Noun] [in/IN] [this/Det] [area/Noun]

Every lexical item that is given a Verb or Noun tag is formed as the headwords/root words. Noun classification words are regarded as entities E, and Verb. Words categorisation are considered as the relations R (among entities). In the example above, the NL question consists of "agent" and "pilgrim" words as an entities, and the word contact is described as a relation among these entities. Another example illustrates the prototype's initial steps in processing questions at Figure 5.

The question consists of a few un-polysemous and polysemous entities. Though this study concentrates on polysemous words, the un-polysemous words are also help in specifying the context during the analysis stages. The words are classified as the following components Verb Phrase (VP), Noun Phrase (NP), and Prepositional Phrase (PP) when each phrase in a sentence is tagged with the respective tag for POS processing. Rule-based chunking receives the labelled phrases and structures the question in a grammatically associated segments for this tasks. Each word is given its chunking and POS tag as revealed in Figure 5 and pseudo-code of the syntactic algorithm is clarified in Algorithm 1.



**Figure 5.** Syntactic processing tasks of a posed NL question via UI of a mobile application QA prototype. These include: Entities/Relations extractions, Chunking, and POS.

#### Algorithm 1 Syntactic\_Analysis\_NLQ

```

1: Load Dt, GPS
2: Input Sentence II Synset [ II ]
3: Variable Identification State IST // Internal Structure
4: Lex // Lexicon
5: SUBF // Chunked Sentence
6: e // Entity
7: POS (Sentence) ← Synset [ II ]
8: for e in Synset [ II ] do
    a. Find_Noun (e, Synset) ← NE[e]
    b. Find_Verb (e, Synset) ← Vr[e]
9: EndFor
10: Find_Concentrate (e, IST) ← SUBF
11: for e in SUBF do
    a. NounPhrase (e, Lex)-SynsetNP [e]
    b. VerbPhrase (e, Lex)-SynsetVP[e]
    c. Preposition (e, SynsetNP)-SynsetP[e]
12: EndFor
13: Return (SynsetNP[e], SynsetVP[e], SynsetP[e]) ← SynInfo

```

#### 4.3. Semantic Analysis

Semantic analysis is mainly related to assigning senses to the grammatical construction that has been established by the syntactical analysis. The result of this process is in the form of a structured semantic representation that enables a system to conduct an appropriate task in its application domain. Consequently, ref. [16] argued that semantic analysis can be considered as an important intermediate stage for natural language understanding (NLU) applications and one of the significant issues in NLP. Semantic analysis engages the mapping of separate lexical items into appropriate objects in the base of knowledge and the generation of precise structures that represent the meaning well when individual lexical items are connected.

Semantic analysis has fundamentally been considered in different domains such as the QA systems. In QA systems, representing target senses is defined by the underlying semantic task. In such a task, the NL question is moved into a proper representation that can be conducted to retrieve the required information. Various accomplishments at the



NLU level are resulting from other domains comprises knowledge representation and information extraction that strive for the same objective which is to have a meaningful representation of sentences in a natural language.

Furthermore, semantic analysis can be classified into deep and shallow semantics analysis. The shallow semantic analysis defines phrases of a sentence with semantic roles relevant to a target lexical item whereas deep-semantic analysis is described by a standard representation of logic languages such as first-order logic. Shallow semantic analysis has been established for the aim of this work to perform three essential tasks. This includes expected answer type, SRL and WSD. The proposed shallow semantic processor gets the benefits of external knowledge resources to do its tasks. Such tasks and the resources of knowledge are highlighted in the next subsections.

#### 4.3.1. Word Sense Disambiguation Module

The most important task in the semantic analysis is word sense disambiguation (WSD) that engages the association of a particular lexical item in a text or discourse with a definition or a meaning; this is not similar to other senses that can be attributed to such a lexical item [16]. The WSD task involves two phases: the determination of all different senses for each lexical item relevant to the text or discourse under consideration and the assignment of each occurrence of a lexical item to an appropriate meaning.

The NL question processing requires the WSD process to determine the most appropriate senses of the polysemous words. The association between “words”, “meaning” and “context” is crucial to define the specific context of the posed NL question which assist in understanding the correct meaning. To achieve the WSD’s major goal, this work integrates the WordNet domain, lexicon domain ontology, and question’s metadata as knowledge resources to support the disambiguation process. WordNet is the most prevalent lexical ambiguity dictionary employed in NL question processing. The output representation form of the syntactic analysis phase is employed as the input form to the semantic analysis processing phase. WordNet domain is utilized in this phase to assign the appropriate senses of the polysemous words based on the NL question’s context. The lexicon domain ontology with the provided data/location information is also utilised to further enhance the disambiguation processing to determine the accurate sense. This ontology has been developed which is comprised of concepts and relations to provide the real-world knowledge of the proposed domain.

#### 4.3.2. Context Knowledge

Determine the correct meaning of the polysemous words depending on the context of the NL question. The context is represented as a group of “features” of the words which includes information about the surrounding words in the NL question. If the question includes terms such as currency, deposit, or cashier this means that the economy domain is the appropriate context domain, in which the polysemous word “bank” refers to a financial institution. Whereas, if surrounding words such as land name, along, river, stand on, located on are detected in the context of NL question this can indicate the geography domain context, and the sense of the polysemous word “bank” is a sloping land. The domain of the context is compared to the target word’s senses in order to disambiguate words by using the external context knowledge such as WordNet. Although WordNet covers a wide range of words and semantic relations, It is limited in words availability and domain-specific terminology in its database [38]. Therefore, the ontology of the domain’s concept beside the WordNet domain is used in this work to assist in lexical disambiguation processing.

For instance, the entity ‘bank’ at the following NL question: “How could a pilgrim deposit money in a bank” is a polysemous lexical item from the received nouns, while the un-polysemous nouns in the NL question such as ‘money’ are utilized to specify the context domain. It is recognised as an un-polysemous noun that is mainly associated with the economy domain. According to WordNet, the noun bank has ten senses from different

context domains categories, six of them are categorised under the economy domain (#2, #3, #4, #6, #8, and #9). This would narrow down the WSD processing to six senses rather than ten senses through the concept ontology.

#### 4.3.3. Date/Location Identifier

The QA system with a meaningful constituent is provided by the question's metadata identified when dealing with context-awareness. KSD QA system is allowed to manage notions such as seasons and location. Since the dates of both the Hajj and the Umrah are different, managing relations in the lexicon domain ontology can be clarified with the help of the identifier. A set of activities make the dates and major locations (Holy places) of the hajj are different from that of the mrah [1]. This can be taken into account with the pilgrim's geographical coordinate and specific landmarks integrated with the ontology concept.

#### 4.3.4. Lexicon Domain Ontology

Ontology has become an essential approach to provide domain-specific knowledge so as to enhance the semantic ability in several domains including the biomedical [39], expert systems [40], intelligent transportation systems [41], forensics [42], and question answering systems [43]. In this study, the ontology is developed to enrich the semantic representation and the relationships among entities for the domain of interest. Concepts are representing entities within the domain like an agent, pilgrims, stone, umrah, hajj, mosques, etc. Verbs represent these relationships question sessions such as offer, request, provide, pray, collect, etc. This association can also be between three entities. In this case, it is represented by a derivation of the main association. For instance, in Figure 6, which shows part of the domain ontology, the relationship 'request In' is derived from the request that connects pilgrims with services and banks. The different meanings which could be assigned to a concept are represented by semantic knowledge that has been extracted from WordNet and interact with other concepts. For example, the sense of bank#2 interacts with entity pilgrim and entity money through the relationship 'deposit-into'. However, the sense of bank#1 interacts with entity pilgrim and entity agent#2 by the relationship 'provided At'.

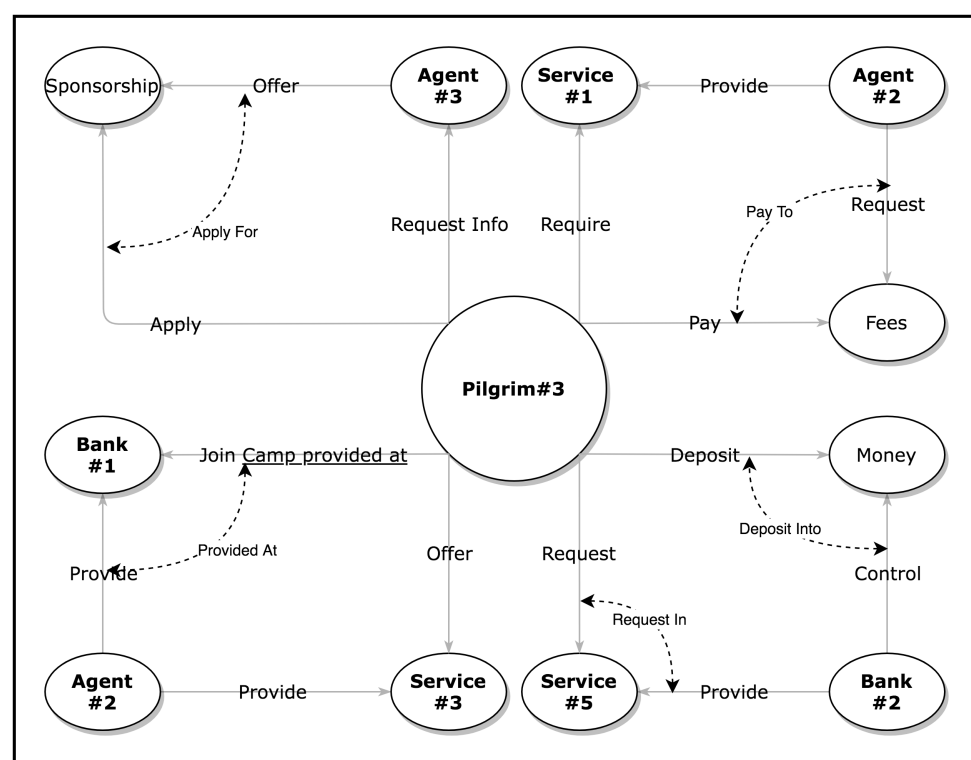
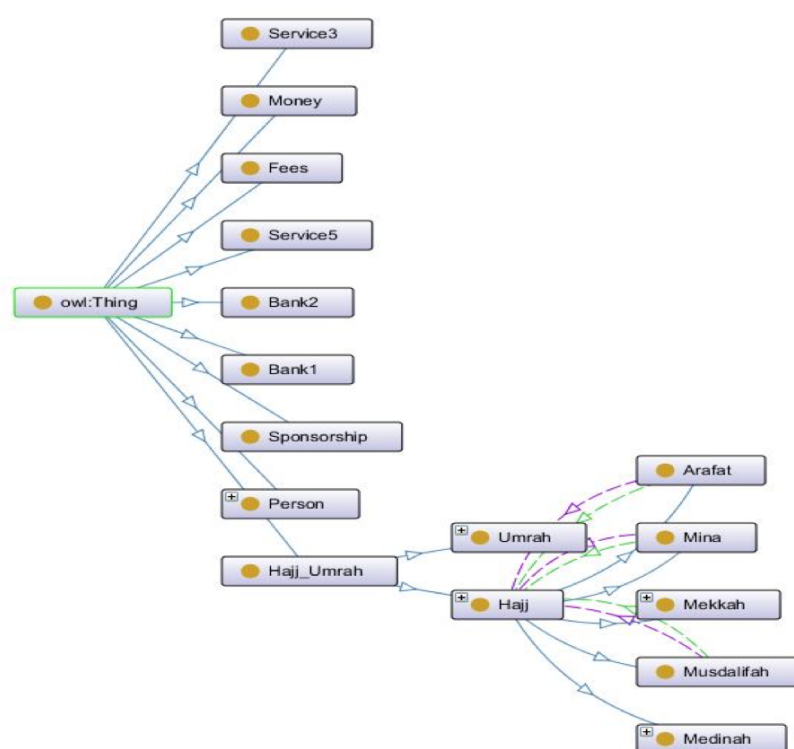


Figure 6. Part of the lexicon domain ontology.

The need for developing ontology to define the set of concepts, entities and their association's relationships to the pilgrimage domain are considered in this study. Therefore, some general asked questions through many websites of the Hajj and Umrah are employed in formulating NL questions. Figure 7 demonstrates a part of the ontology that has been developed in the context of a pilgrimage domain. The ontology is structured in the form of a graph, which is comprised of entities, relationships and arrows that demonstrate an association among them.



**Figure 7.** Part of ontology development in protégé.

#### 4.3.5. Semantic Role Labelling

The automatic labelling of semantic roles serves to determine the meaning of the received question. This is achieved by identifying the classification of arguments and predicate of the question. Accordingly, SRL structures the semantic representation in the form of predicate–argument, which supports the processing of answer matching. This SRL structure involves the representation of words attached with related semantic senses. The predicate–argument structure associated with the QA system denotes  $P(SRLs)$ , where  $P$  represents the predicate, and the SRLs is a group of semantic role labels. The initial syntactic analysis is crucial, which assists to determine arguments associated with the predicate of the NL question [14]. In this study, nouns are considered as arguments associated with a verb type predicate of the NL question. Syntactic prepositions are either classified as spatial relationships, e.g., in, on, into, onto or instrumental relationships, e.g., with and by. For instance, the classification of entities in the previous NL question includes deposit classified as Pred, pilgrim as Arg0, money as Arg1, and bank as Arg-Loc. Algorithm 2 demonstrates the pseudo code of the semantic analysis algorithm. Figure 8 illustrates the semantic analysis processing which has been performed to disambiguate the polysemous word bank in the previous NL question. The process of the lexical item disambiguation is illustrated in this algorithm, and then a final semantic representation is assigned utilizing the SRL method.

#### 4.3.6. Expected Answer Type

Classifying the answer type of the question is the final phase in semantic analysis processing. It imposes restrictions on the possible answers, which is a significant role to detect the specific answer to the question. For instance, in the following question "When can I apply for the Hajj this year?", the potential candidate categorization of the answer type is "time". To achieve this, the expected answer type classification rules are provided for this phase.

---

#### Algorithm 2 Semantic\_Analysis\_NL Question (Syninfo)

---

```

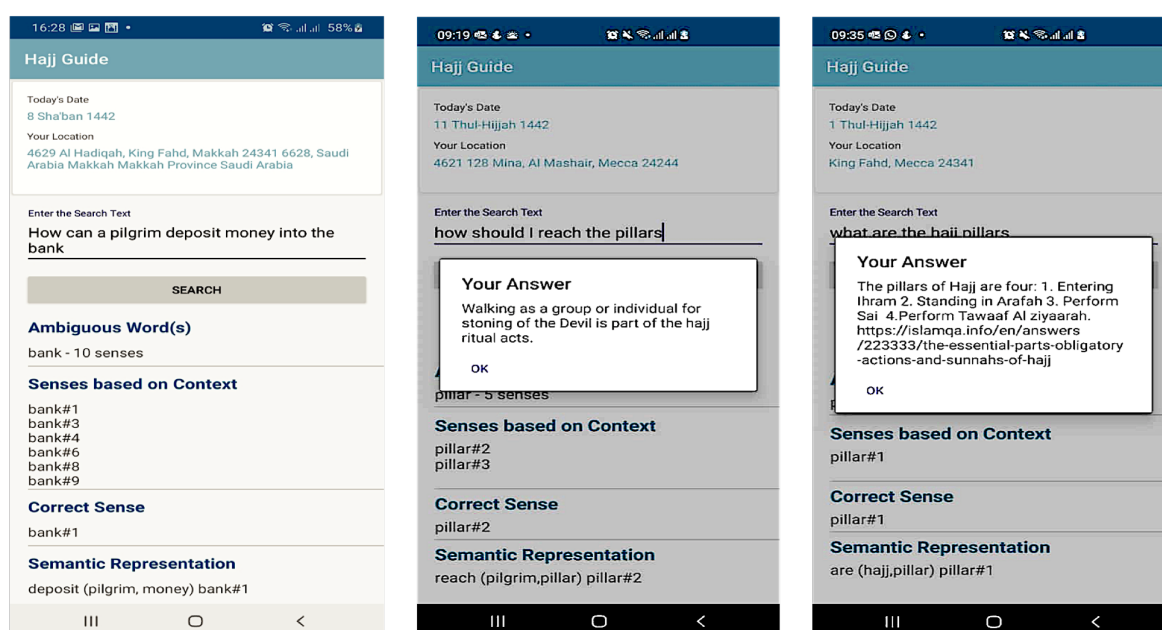
1: Load Dt, GPS // Date and Location Coordinates
2: Input Syninfo [N]
3: Define Umrah_DateRange
4: Define Umrah_Location
5: Define Hajj_DateRange
6: Define Hajj_Location
7: Variable Identification
8: Syninfo: Syntactic Information from the shallow syntactic analyzer
9: DO: Domain Ontology
10: Synset : Syntactic set after the remover of irrelevant information from Syninfo
11: CS : Correct Sense
12: Entity e
13: for each entity e in Syninfo do
    a. Non_Relevant_Concept (Syninfo) → Synset[e]
    b. Find_Ambiguity (e, Synset) → AE[e]
    c. Find_Context (e, Synset) → Cr[e]
14: If (AE[e] == 1) then SenseAE[e]
15: Elseif (Dt.Umrah_DateRange() and GPS.Umrah_Location()) → Sense(AE[e], Umrah)
16: Map (Dt, GPS) → UmrahDO
17: CSUmrah → Sense(UmrahDO)
18: Return Sense(AE[e], CSUmrah)
19: Elseif (Dt.Hajj_DateRange() and GPS.Hajj_Location()) → Sense(AE[e], Hajj)
20: Map (Dt, GPS) → UmrahDO
21: CSUmrah → Sense(UmrahDO)
22: Return Sense(AE[e], CSUmrah)
23: Endif
24: EndFor
25: DO : SenseRelation
26: If Sense_In VP then Sense Predicate // Verb Phrase (VP)
27: If Sense_Pred NP then Predicate Subject // Noun Phrase (NP)
28: If Sense_Fol NP then Predicate Object
29: Sense_InPP then Sense – Location or Instrument // Predicate Phrase (PP)
30: EndIf

```

---

#### 4.4. Answer Processing Module

The constituent answer module aims to process the matching between the obtained semantic structure form of the NL question (NL question\_Sem) and the possible answer (A\_Sem). In this study, the possible answers are structured in the form of a predicate-argument in a database. A rules-based matching model is developed to support the answer processing module to detect the candidate answer. Accordingly, the candidate answer (A\_Sem) is detected for (NL question\_Sem) by the use of automatic matching on the basis of the potential answer classification types. The candidate answer matching is processed according to the classification of answer type. The candidate answers are sent back to the user corresponding to the posted NL question.



**Figure 8.** Semantic analysis processing tasks of posed NL questions via UI of a mobile application QA prototype. These include: the polysemous word, number of senses based on the context of the NL question, the correct sense identified, and the semantic representation. The figures also include two NL questions contain “pillars” as a polysemous word. Although “pillars” in this context has different senses, the system were capable to determined the correct sense for each posed question and returned the correct answer accordingly.

## 5. Evaluation

This section discusses the experiments evaluation details of the proposed method. The comparison between one system and another is difficult due to the various knowledge resources adopted, sense inventories, experiment conditions, datasets, environment, and domain [22]. This has posed a difficult task since there are no existing comparable systems and datasets in terms of the proposed domain, where most of the existing related works aim to provide a direct answer to questions and not consider the problem of ambiguity associated with NL questions [1,19]. Thus, the evaluation process goes through two types of experiment procedures for performance evaluation. Although evaluating the proposed approach through one of these types is enough [5]. In vitro experiment is to test and evaluate the performance of KSD method for WSD in comparison to the the simplified Lesk extended variant SE-Lesk [44], and the most frequent sense (MFS) defined by WordNet [45,46]. In vivo testing is to validate and evaluate the effectiveness of the KSD approach into an overall QA system’s performance in the context of the proposed domain. To achieve these evaluations, the improvement results would be counted to evaluate the performance of the methods in terms of accuracy [11,46].

### 5.1. Experimental Settings

Android mobile operating system was selected as the platform of choice due to the nature of the proposed pilgrimage domain and the fact that it has a more user base. Android Studio 4.2 IDE with a Java platform (JDK 8). Besides, my SQLite and Android Programming (Java and XML) as the database environment are used as instruments to create the GUI application, to code and implement the mobile application prototype elements. The main external sources used in the experiments are the lexical database WordNet 3.1 and the proposed ontological information. Protégé software (V 5.5.0) was used for ontology development. In order to test and evaluate our approach, 91 NL questions were extracted from the pilgrimage frequently asked questions, that were collected from several related websites, predominantly from the pilgrimage official websites for the test dataset.

To set up the experiment environment, a lexicon containing over 400 words were conducted for the experiment. This contains all the words that feature in our NL questions such as service, agent, camp, bank, area. These words were classified into the following types nouns, verbs, pronouns, prepositions, adverbs, adjectives, determiners and auxiliary verbs. A word is polysemous; belonging to more than one category, e.g., contact may be verbs or nouns. Some set of rules have been put in place to determine the right one. It is crucial to categorize words in their correct part of speech; otherwise, this affects further phases. In addition to POS, the chunking engine was also built to form phrases. The trio of a noun, verb and prepositional is considered in this work.

The context knowledge base of 400 words is developed where every single term is labelled with the appropriate lexical senses and associated contexts. These comprise the most commonly used terms in the context of the pilgrimage domain. Accordingly, the ontology was designed for the domain, then the actual construction of the ontology concepts of the Hajj and Umrah services were identified and developed. These include pilgrim, camp, bank, agent, etc. Additionally, over 250 relations that link these concepts were identified and developed in the ontology. The test dataset consists of 97 polysemous words from a set of 91 NL questions. Table 1 shows the sample of the posed questions, polysemous words, and status based on the KSD method.

**Table 1.** Sample of the posed NL questions (Processed), number of polysemous word (No.PW), resolved, and answered.

SN	Processed	No.PW	Resolved	Answered
1	How can a pilgrim deposit money into the <b>bank</b> ?	1	Yes	Yes
2	What happen if a pilgrim collect pebbles at the <b>bank</b> of muzdalifah?	1	Yes	Yes
3	What kind of <b>facilities</b> does the <b>agent</b> offer in the <b>camp</b> ?	3	No	No
4	What are the <b>dates</b> of the Hajj?	1	Yes	Yes
5	What are the <b>dates</b> available from Almadinah?	1	Yes	Yes
6	What are the essential <b>parts</b> of Hajj?	1	Yes	Yes
7	What are the obligatory <b>actions</b> of Hajj?	1	Yes	Yes
8	When should the pilgrims join the <b>camp</b> ?	1	Yes	Yes
9	How can a pilgrim <b>contact</b> an <b>agent</b> at the <b>camp</b> ?	3	No	No
10	What are the important <b>dates</b> in Hajj?	1	Yes	Yes
11	What are the benefits of <b>dates</b> ?	1	Yes	Yes
12	How can a pilgrim explore important places in the <b>area</b> ?	1	Yes	Yes
13	Where can I find <b>scholar</b> in the <b>area</b> ?	2	Yes	Yes
14	Where can I seek advise with personal <b>service</b> concern? ?	1	Yes	Yes
16	When should a pilgrim manage their <b>trip</b> to Mina?	1	Yes	Yes
17	What are the hajj <b>pillars</b> ?	1	Yes	Yes
18	How should I reach the <b>pillars</b> ?	1	Yes	Yes

## 5.2. Evaluation Results

The experiment was conducted where the three methods, KSD, SE-Lesk, and MFS, were tested on the same condition and test dataset. Each method takes questions as the input and outputs the appropriate sense for each ambiguous term found in each question. The test dataset includes several types of NL questions ranging across different Wh-questions. The results are measured based on the sense disambiguation performance of the three implemented methods. Accuracy is the most essential and simple evaluation metric for WSD. It calculates the number of correctly disambiguated words divide by the total number of polysemous words [5,46,47]. The results of the evaluation show that our KSD method achieves 84.5% while MFS achieves 79.3% and 65.9% achieved by the



SE-Lesk algorithm based on the number of correctly disambiguated words (CD column) as illustrated in Table 2. The experiment also shows that in some cases, the processors were unable to determine the correct senses of the polysemous words. This occurs when a short question includes a few ambiguous words that make it challenging to decide the relevant context. Solving ambiguity issues is highly dependent on the local context, hence, the appearance of more surrounding words may improve the performance.

**Table 2.** Results of the evaluation includes method, number of polysemous word (No.PW), correctly disambiguated (CD) word, and accuracy.

Method	No.PW	CD	Accuracy
KSD	97	82	84.5%
MFS	97	77	79.3%
SE-Lesk	97	64	65.9%

Additionally, we carried out a further experiment to measure the impact of our KSD approach on a QA system performance in the pilgrimage domain. In this experiment, we evaluate our approach in comparison to the MFS baseline. For this experiment, another mobile QA prototype has been developed based on MFS. The performance of KSD was compared with MFS in processed correct answers based on the correct semantic representation for the posed 91 questions through the mobile QA prototypes in the same condition, locations and test dataset.

The experimental results illustrated in Table 3 include the processed NL questions for each system that were correctly answered (CA column). Accuracy is the standard evaluation metric to test and evaluate the performance of QA systems. It is calculated by dividing the number of correctly answered questions by the total number of questions in the test dataset [9]. The experiment results indicate the affect of the KSD-based QA prototype in returning correct responses to the posed questions, which achieve 6.6% improvement in comparison to the MFS-based QA prototype in the context of the pilgrimage domain.

**Table 3.** Results of the evaluation includes system, number of questions (NLQs), correctly answered (CA) questions and accuracy.

System	NLQs	CA	Accuracy
KSD-based QA system	91	80	87.9%
MFS-based QA system	91	74	81.3%

## 6. Discussion

The proposed KSD method aims at enhancing the accuracy of the lexical sense disambiguation process in NL questions. This method incorporates question's metadata, context knowledge, and the domain ontological information, into a shallow NLP method. It supports the capabilities to determine the intended meaning of questions posed in NL to effectively answer corresponding questions in a specific domain. This work investigates the capabilities of the proposed solution to assist pilgrims in expressive queries to obtain concise and accurate information via a QA application through their smartphones. Although the ontological information supports lexical disambiguation and reasoning capabilities by providing unambiguous specific knowledge representation in different granularity levels, the incomplete coverage of the domain ontology may negatively impact our method's efficiency. This failure was observed in some cases during experiments in comparison to SE-Lesk and MFS where the proposed method was unable to process the appropriate sense due to the absence of related entities and relations on domain ontology. Another major difference in the results was the significant performance achieved by MFS in handling multiple ambiguous words in short questions during both experiments compared to KSD and SE-Lesk. This would be considered for the future next steps besides the

possibility of available comprehensive reliable ontology or the automatic construction with consistent coverage expandability of the domain ontology is worth investigating.

## 7. Conclusions and Future Works

In this paper, we propose a novel knowledge-based sense disambiguation KSD method. KSD aims to determine the intended meaning and assign the correct sense to the polysemous words that appear in NL questions to enhance the QA system performance. The proposed approach integrates dictionary-based and artificial intelligence-based methods which categorised it as a knowledge-based approach. Our approach integrates multiple types of knowledge: question's metadata, context, and the domain ontology into the shallow-based analysis. The context knowledge is assisted by the WordNet 3.1. The domain ontology is designed to represent rich and complex real knowledge in the form of entities and their relations to describe concepts of the domain. However, the incompleteness of the lexicon domain ontology may lead to low performance. The results of the evaluation show that the proposed KSD approach achieves comparable and better performance compared to SE-Lesk and MFS baselines in the pilgrimage domain. In the future, we are planning to investigate the advancement of these methods within an expandable framework by incorporating an explainable knowledge base and feature learning model for word embedding such as the Word2Vec model to learn from domain-related data and predict wider relations and entities. The integration of these models involves the expansion of a new framework. An obvious future avenue is the development of a tool for the pilgrimage QA application based on the advancement method to accomplish various complex question types effectively.

**Author Contributions:** Conceptualization, A.A.; Data curation, A.S.; Formal analysis, A.A.; Investigation, A.A.; Methodology, A.A. and A.S.; Project administration, A.A.; Resources, A.A. and A.S.; Software, A.A.; Supervision, A.S.; Writing – original draft, A.A.; Writing – review & editing, A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not Applicable, the study does not report any data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Arbaeen, A.; Shah, A. Ontology-Based Approach to Semantically Enhanced Question Answering for Closed Domain: A Review. *Information* **2021**, *12*, 200.
2. Al-Harbi, O.; Jusoh, S.; Norwawi, N.M. Lexical disambiguation in natural language questions (NLQs). *arXiv* **2017**, arXiv:1709.09250.
3. Ojokoh, B.; Adebisi, E. A Review of Question Answering Systems. *J. Web Eng.* **2018**, *17*, 717–758.
4. Pundge, A.M.; Khillare, S.; Mahender, C.N. Question Answering System, Approaches and Techniques: A Review. *Int. J. Comput. Appl.* **2016**, *141*, 0975–8887.
5. Navigli, R. Word sense disambiguation: A survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 10.
6. Höffner, K.; Walter, S.; Marx, E.; Usbeck, R.; Lehmann, J.; Ngonga Ngomo, A.C. Survey on challenges of question answering in the semantic web. *Semant. Web* **2017**, *8*, 895–920.
7. Correa, E.A., Jr.; Lopes, A.A.; Amancio, D.R. Word sense disambiguation: A complex network approach. *Inf. Sci.* **2018**, *442*, 103–113.
8. Wang, Y.; Wang, M.; Fujita, H. Word sense disambiguation: A comprehensive knowledge exploitation framework. *Knowl.-Based Syst.* **2020**, *190*, 105030.
9. Jabalameli, M.; Nematbakhsh, M.; Zaeri, A. Ontology-lexicon-based question answering over linked data. *ETRI J.* **2020**, *42*, 239–246.
10. Al Fawareh, H.M.K. Resolving Ambiguity in Entity and Fact Extraction through a Hybrid Approach. Ph.D. Thesis, Universiti Utara Malaysia, Bukit Kayu Hitam, Malaysia, 2010.

11. Raganato, A.; Camacho-Collados, J.; Navigli, R. Word sense disambiguation: A unified evaluation framework and empirical comparison. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain, 3–7 April 2017; Volume 1, Long Papers; pp. 99–110.
12. Navigli, R. Natural Language Understanding: Instructions for (Present and Future) Use. In Proceedings of the IJCAI, Stockholm, Sweden, 13–19 July 2018; Volume 18, pp. 5697–5702.
13. Mohammed, S.; Shi, P.; Lin, J. Strong baselines for simple question answering over knowledge graphs with and without neural networks. *arXiv* **2017**, arXiv:1712.01969.
14. Pillai, L.R.; Veena, G.; Gupta, D. A combined approach using semantic role labelling and word sense disambiguation for question generation and answer extraction. In Proceedings of the 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAIECC), Bangalore, India, 9–10 February 2018; pp. 1–6.
15. Aouicha, M.B.; Taieb, M.A.H.; Marai, H.I. WSD-TIC: Word Sense Disambiguation Using Taxonomic Information Content. In Proceedings of the International Conference on Computational Collective Intelligence, Halkidiki, Greece, 28–30 September 2016; Springer: Cham, Switzerland, 2016; pp. 131–142.
16. Mennes, J.; van Gulik, S.v.d.W. A critical analysis and explication of word sense disambiguation as approached by natural language processing. *Lingua* **2020**, *243*, 102896.
17. White, R.W.; Richardson, M.; Yih, W.t. Questions vs. queries in informational search tasks. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; ACM: New York, NY, USA, 2015; pp. 135–136.
18. del Carmen Rodriguez-Hernández, M.; Ilarri, S.; Trillo-Lado, R.; Guerra, F. Towards keyword-based pull recommendation systems. In Proceedings of the ICEIS 2016, Roma, Italy, 25–28 April 2016; p. 207.
19. Khan, E.A.; Shambour, M.K.Y. An analytical study of mobile applications for Hajj and Umrah services. *Appl. Comput. Inform.* **2018**, *14*, 37–47.
20. Arbaeen, A.; Shah, A. Natural Language Processing based Question Answering Techniques: A Survey. In Proceedings of the 2020 IEEE 7th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Kuala Lumpur, Malaysia, 18–20 December 2020; pp. 1–8.
21. Rodrigo, A.; Penas, A. A study about the future evaluation of Question-Answering systems. *Knowl.-Based Syst.* **2017**, *137*, 83–93.
22. Chaplot, D.S.; Salakhutdinov, R. Knowledge-based word sense disambiguation using topic models. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
23. Guo, J.; Fan, Y.; Pang, L.; Yang, L.; Ai, Q.; Zamani, H.; Wu, C.; Croft, W.B.; Cheng, X. A deep look into neural ranking models for information retrieval. *Inf. Process. Manag.* **2020**, *57*, 102067.
24. Wu, Y.; Hori, C.; Kashioka, H.; Kawai, H. Leveraging social Q&A collections for improving complex question answering. *Comput. Speech Lang.* **2015**, *29*, 1–19.
25. Cui, W.; Xiao, Y.; Wang, H.; Song, Y.; Hwang, S.w.; Wang, W. KBQA: Learning question answering over QA corpora and knowledge bases. *arXiv* **2019**, arXiv:1903.02419.
26. Figueroa, A.; Neumann, G. Context-aware semantic classification of search queries for browsing community question-answering archives. *Knowl.-Based Syst.* **2016**, *96*, 1–13.
27. Pechsiri, C.; Piriyaikul, R. Developing a Why-How Question Answering system on community web boards with a causality graph including procedural knowledge. *Inf. Process. Agric.* **2016**, *3*, 36–53.
28. Khodadi, I.; Abadeh, M.S. Genetic programming-based feature learning for question answering. *Inf. Process. Manag.* **2016**, *52*, 340–357.
29. Chali, Y.; Hasan, S.A.; Mojahid, M. A reinforcement learning formulation to the complex question answering problem. *Inf. Process. Manag.* **2015**, *51*, 252–272.
30. Yang, M.C.; Lee, D.G.; Park, S.Y.; Rim, H.C. Knowledge-based question answering using the semantic embedding space. *Expert Syst. Appl.* **2015**, *42*, 9086–9104.
31. Hao, Y.; Zhang, Y.; Liu, K.; He, S.; Liu, Z.; Wu, H.; Zhao, J. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017; Volume 1: Long Papers; pp. 221–231.
32. Sun, H.; Dhingra, B.; Zaheer, M.; Mazaitis, K.; Salakhutdinov, R.; Cohen, W.W. Open domain question answering using early fusion of knowledge bases and text. *arXiv* **2018**, arXiv:1809.00782.
33. Saloot, M.A.; Idris, N.; Mahmud, R.; Ja’afar, S.; Thorleuchter, D.; Gani, A. Hadith data mining and classification: A comparative analysis. *Artif. Intell. Rev.* **2016**, *46*, 113–128.
34. Sulaiman, S.; Mohamed, H.; Arshad, M.R.M.; Yusof, U.K. Hajj-QAES: A knowledge-based expert system to support hajj pilgrims in decision making. In Proceedings of the 2009 International Conference on Computer Technology and Development, Kota Kinabalu, Malaysia, 13–15 November 2009; Volume 1, pp. 442–446.
35. Sharef, N.M.; Murad, M.A.; Mustapha, A.; Shishechi, S. Semantic question answering of umrah pilgrims to enable self-guided education. In Proceedings of the 2013 13th International Conference on Intelligent Systems Design and Applications, Salangor, Malaysia, 8–10 December 2013; pp. 141–146.
36. Mohamed, H.H.; Arshad, M.R.H.M.; Azmi, M.D. M-HAJJ DSS: A mobile decision support system for Hajj pilgrims. In Proceedings of the 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 15–17 August 2016; pp. 132–136.

37. Abdelazeez, M.A.; Shaout, A. Pilgrim Communication Using Mobile Phones. *J. Image Graph.* **2016**, *4*, doi:10.18178/joig.4.1.59-62.
38. Dhungana, U.R. Polywordnet: A Word Sense Disambiguation Specific Wordnet of Polysemy Words. Ph.D. Thesis, Tribhuvan University, Kirtipur, Nepal, 2017.
39. Alobaidi, M.; Malik, K.M.; Sabra, S. Linked open data-based framework for automatic biomedical ontology generation. *BMC Bioinform.* **2018**, *19*, 319.
40. Nogueira, T.P.; Braga, R.B.; de Oliveira, C.T.; Martin, H. FrameSTEP: A framework for annotating semantic trajectories based on episodes. *Expert Syst. Appl.* **2018**, *92*, 533–545.
41. Ali, F.; El-Sappagh, S.; Kwak, D. Fuzzy Ontology and LSTM-Based Text Mining: A Transportation Network Monitoring System for Assisting Travel. *Sensors* **2019**, *19*, 234.
42. Wimmer, H.; Chen, L.; Narock, T. Ontologies and the Semantic Web for Digital Investigation Tool Selection. *J. Digit. Forensics Secur. Law* **2018**, *13*, 6.
43. Jiang, S.; Wu, W.; Tomita, N.; Gao, C.; Hassanpour, S. Multi-Ontology Refined Embeddings (MORE): A Hybrid Multi-Ontology and Corpus-based Semantic Representation for Biomedical Concepts. *arXiv* **2020**, arXiv:2004.06555.
44. Banerjee, S.; Pedersen, T. An adapted Lesk algorithm for word sense disambiguation using WordNet. In Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics, Mexico City, Mexico, 17–23 February 2002; Springer: Berlin/Heidelberg, Germany, 2002; pp. 136–145.
45. Agirre, E.; Edmonds, P. *Word Sense Disambiguation: Algorithms and Applications*; Springer Science & Business Media: New York, NY, USA, 2007; Volume 33.
46. Oele, D.; Van Noord, G. Distributional lesk: Effective knowledge-based word sense disambiguation. In Proceedings of the IWCS 2017—12th International Conference on Computational Semantics, Montpellier, France, 19–22 September 2017; Short Papers.
47. Badugu, S.; Manivannan, R. A study on different closed domain question answering approaches. *Int. J. Speech Technol.* **2020**, *23*, 315–325.