International Journal of Digital Information and Wireless Communications (IJDIWC) 4(2): 236-247
The Society of Digital Information and Wireless Communications, 2014 (ISSN: 2225-658X)

Designing and Implementing Bi-Lingual Mobile Dictionary to be used in Machine Translation

Hassanin M. Al-Barhamtoshy
Faculty of Computing and Information Technology
King Abdulaziz University (KAU)
Jeddah, Saudi Arabia
hassanin@kau.edu.sa

Fatimah M. Mujallid
Computer Science Dept., Faculty of Computing
King Abdulaziz University (KAU)
Jeddah, Saudi Arabia
f.mujallid@gmail.com

ABSTRACT

This paper describes the multistage process for building Arabic WordNet (ArWn) to be used in mobile device. The goal of this paper is how to create corpus, starting with selecting an annotation task, designing the data with the annotation process, and finally evaluating the results for a particular goal. Therefore, the paper presents designing and implementing bi-lingual lexicon to be used in machine translation and language processing.

Consequently, the paper takes into consideration language characteristics in both directions Arabic and English. The proposed system is based on WordNet lexical database with a semantic and commonsense knowledge. The proposed dictionary will be implemented for mobile devices, therefore; the cloud computing will be used in this implementation. Consequently, SQL Azure will be used to solve scalability, and interoperability of mobile users and other methods have been used for both Arabic and English languages. So, the SQL Azure will be used as the cloud database to solve both the scalability in the data with scale terabytes of data to millions of mobile users and the interoperability challenges. The system dictionary is developed and tested in Android mobile platform. Experimental results show that the proposed system has two versions- at work; offline and online. The online approach uses the mobiles computing in the cloud system to reduce the storage complexity of the mobile. Real time test will be used in order to evaluate the system access and respond times to display results.

KEYWORDS

MT, dictionary, Arabic, NLP, lexical, and commonsense.

1 INTRODUCTION

Machine Translation (MT) is an important area of Natural Language Processing (NLP) applications and technologies in this domain are highly required. Machine Translation applications translate source language text (SL) into target language text (TL) [1] [2]. Multilingual chat applications, emails translation, and real-time translation of web sites are typical examples of machine translation.

In multilingual applications, machine translation (MT) is an essential component, and it is highly-demanded technology in its own right. Multilingual chatting, talking translators, and real-time translation of emails and websites are some examples of the modern commercial applications of machine translation.

Typically, dictionaries have been used in human translation, and have also been used for dictionary-based machine translation.

The main challenges that machine translation systems encounter can be divided into two categories: missing words, translation variants, and deciding on whether or not to translate a name (or part of it).

Conventionally, semantic resources and lexicons have been used as core components for building different applications in NLP. Recently, researchers and developers have been using lexical databases in NLP applications [3] [4]. Semantic resources can be performed from lexical database within several domains. Morphology, syntactic and semantic features are needed to drive lexical items of individual lexical items. Bilingual and multilingual dictionaries are lexical databases and they are depending on the type of languages that they are involved [5]. Semantic, commonsense knowledge's and more semantic information about specific word can be produced from lexical database. One of the most widely known commonsense knowledge bases is WordNet[1] [2] [6] [7].

[1] Wikipedia lexical resource: http://en.wikipedia.org/wiki/Lexical_resource
[2] What is WordNet? http://WordNet.princeton.edu/WordNet/

Arabic language is one of the most spoken language in a group called semitic languages, 422 people around the world speak it which considered to be one of most considered and distributed language around the globe [8] [9] [10] [11] [12]. The Arabic language is ranked sixth of the most ten impact languages, with an estimated 186 million native speakers. In 2010 [12] the number of Arabic native speakers increased to 239 million people and the ranked of Arabic in the list rose to the fifth[3]. Arabic speakers are increasing and Arabic language is expanding in the world, therefore number of Arabic documents and articles are increased. This shows the importance of the Arabic Language in the world.

Currently, linguistic and lexical resources for Arabic language are growing but still they are few, especially efforts for mobile devices. However, the last decade has known a number of attempts aiming at offering electronic resources for the Arabic NLP community. One of the attempts is the Arabic WordNet [12] [13] [14] [15] [16] project which the objective was to construct and develop a freely available lexical database for standard Arabic. Arabic WordNet has very low coverage and limited words.

Nowadays, people use their mobile for many purposes and most of the users have replaced computers' desktops and laptops with them. By 2012 there were about 6 billion mobile users in the world[3]. This big number shows what the future will be; mobile computing. There are successful attempts to build English smart mobile dictionary but there are reared in Arabic language. The need for an Arabic lexical database mobile application has led to the creation of mobile dictionary system. This paper presents to design and implement bilingual (Arabic-English) mobile dictionary using WordNet as lexical database.

In this paper, key terminology and formulations used throughout this paper will be introduced. Section 2 gives an overview in all the relevant areas most notably the related work upon this work is founded. Section 3 describes the mobile dictionary framework, so, the system architecture will be presented and illustrated. In section 3, also, the system database has been explained and the system workflow is introduced. Section 4 will discuss evaluation and system performance. We also examine the evaluation procedure undertaken in this paper, and the difficulties that arise with non-standard evaluation methodologies that are often used in the translation area. And last Section gives the conclusion, and future works.

2 LITERATURE REVIEW

Many attempts have been done, to create a dictionary based in WordNet in different languages. The first attempt was Princeton WordNet (PWN)[4,5,6]. The Princeton WordNet has been developed in 1985; it is large lexical database for English language. The words' structure of the PWN is located according to conceptual similarity with other words; to represent semantic dictionary. Therefore, the words that have the same meaning are grouped together in a group called Synset and the words are classified into four parts of speech (POS): nouns, verbs, adjectives and adverbs. Synsets are composed from semantic and lexical relations.

After PWN appearance, many attempts have been emerged to create WordNets for other languages, Euro WordNet (EWN) was a step towards multilingual WordNet [17] [18]. The first release of the EWN was for Dutch, Spanish, Italian, German, French, Czech and Estonian. The structure for each language in EWN is like as PWN. All the EWN languages are connected by an inter-lingual- index (ILI) which connects the Synsets that are the same in different languages. Another project called Balkanian WordNet (BalkaNet) has been created, followed EWN and added more languages such as Bulgarian, Greek, Rumanian, Serbian, and Turkish.

After that, Global WordNet Associations (GWNA)[5] [22] has been created in 2000; and many other languages have been built such as China, Hindi[6] and Korean.

For Arabic language efforts, there is Arabic WordNet (AWN) which is a multilingual lexical

[3] http://newsfeed.time.com/2013/03/25/more-people-have-cell-phones-than-toilets-u-n-study-shows/

[4] "Euro WordNet," (Wikipedia, the free encyclopedia), http://en.wikipedia.org/wiki/EuroWordNet

[5] "The Global WordNet Association," (The Global WordNet Association), http://www.globalWordNet.org/ "Euro WordNet"

[6] Hindi WordNet: http://www.cfilt.iib.ac.in/WordNet/webhwn/

database and it is linked to PWN using ontology inter-lingual mechanism. The structure of AWN consists of four entity types: item, word, form and link. An item has information about the synsets, ontology classes and instances. A word has information about word senses. A form represents a root or is plural form derivation. A link is used to connect two items, and also it connects a PWN synset to an AWN synset. Another WordNet created for Arabic is a master thesis written in 2010 [20]. This thesis presents easy to use Arabic interface WordNet dictionary which is developed as the way the EWN has been developed [21]. This is monolingual dictionary for Arabic language and is not connected to EWN or PWN although it is built following them [21].

All these previous studies were built to work on desktop applications. However there are few attempts to build lexical database on mobile platforms based on lexical knowledge and commonsense. One of these attempts is creating WordNet mobile-base to work with PWN for the Pocket PC platform (Windows Mobile), they called it WordNetCE [22]. Also there is smart phone version (WordNetCE-SP) [23] [24].

Another success attempts is the Dubsar project [24] which is a simple web-based dictionary application based on PWN. Dubsar is a work in progress; it is available for free worldwide on the iTunes App Store for many of mobile devices. Also it is available in the Android Market for free worldwide.

There are other non free dictionaries and thesaurus based on PWN for mobile platform such as English WordNet dictionary by Konstantin Klyatskin[7], Advanced English Dictionary and Thesaurus by Mobile System Company[8], LinkedWord Dictionary & Thesaurus by Taisuke Fujita and Blends by Leonel Martins[9].

From this literature review, the authors can observe that there are no attempts to create an Arabic dictionary for mobile platforms by using lexical database. So the goal is to conduct a dictionary which is organized by meaning and has common- sense, semantic and lexical relations and form a network of meaningfully related terms and concepts. Also it composed of most common and concise English/Arabic words and corresponding explanations and it has quick and dynamic search and works offline and online.

3. FRAMEWORK FORMULATION

To enable consistent explanations of the systems throughout this paper, we define a framework for the proposed translation model and the system that follow this model. The formulation for the translation process, apply primarily to generative transformation method of bilingual translation corpus and evaluation applies to generative and extractive translation approaches.

Therefore, a framework for translation model will be defined in this section. Bilingual dictionary, lexicon and corpus will be used to generate and extract translation approaches. The generative translation process uses two stages: training and generative stages. The two stages running on a bilingual corpus; BC = { (D_S , D_T) }; and the generation stage produces one or more word WT for each source word WS, see Figure 1.

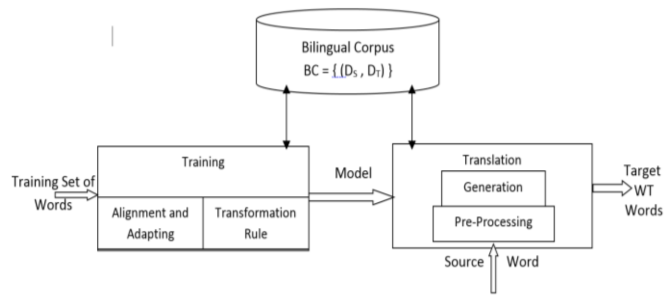

Figure 1. Translation Model Framework

The training stage of the proposed model is composed from three sun-modules: alignment between source and target, segmentation using graphemes or phonemes (in case of speech); and transformation rule to generate the model that built in the bilingual corpus.

Statistical machine translation (SMT) is used in alignment, such SMT model can be considered as a function of faith-fullness to the source language, and fluency in target language [2] [3]. The fundamental model of the SMT is defined based

7 http://filedir.com/company/konstantin-klyatskin/
8 http://appworld.blackberry.com/webstore/content/314/?countrycode=SA&lang=en
9 https://itunes.apple.com/us/app/linkedword-dictionary-thesaurus/id326103984?mt=8

on faith fullness (translation model) and fluency (language model) as the following:

$$P(S, T) = \text{argmax}_T\, P(S|T)\, P(T) \quad (1)$$

Where S and T represent the sentences (words) in source and target languages; P (S | T) represents translation model; and P(T) indicates target language model. Therefore, we need a decoder that, given the sentence (or word) S, produces the most probable sentence (or word) T.

3.1. Alignment

The word alignment is important as a component in machine translation, especially in Statistical machine translation, and, it is defined as it is a mapping between the words of pair sentences that are a translation of each other. Also, alignments can be one-to-one, one-to- many and many-to-many relations. However, it is possible to generate multiple target variants for a word where some translators may add extra vowels to make variants easier to understand.

3.2 Transformation Rules

A transformation rule can be defines as S ➔ (T , p); where S is the source word; T is the target word; and p is the probability of translating S to T. Consequently, for any S that contains n rules, so:

$$S \rightarrow (T_k, p_k) \text{ such that } \sum p_k = 1 \quad (2)$$

Another transformation rule to represent model M is defined as; the model M takes source word S and outputs list of tuples with (Tj , Pj) as its elements. So;

$$S \rightarrow (T_j, P_j) \quad (3)$$

Where; T_j represents tuple with j^{th} rule of the source words generated with j^{th} highest probability P_j.

3.3 Bilingual Corpus

A bilingual corpus BC is defined as transformation pairs { (D_S , D_T) }, where D_s = w_{s1}, w_{s2}, … w_{sl}; and D_T = { W_{Tk}} and W_{Tk} = w_{t1}, w_{t2}, … w_{tm} ; w_{si} is a word in the source language, w_{tj} is word in the target language. Such corpus will be implemented as computerized resources.

3.4 Evaluation Measures

One of the evaluation measures for machine translation is word accuracy. Other metrics are also used in the literature of [3]. Such evaluation schemes can be classified into two categories: single-variant and multi-variant metrics.

3.4.1 Single Variant

Word accuracy is –one of the standard- used to measure evaluation of machine translation. Therefore, word accuracy or transformation accuracy (A) can be calculated as the following formula:

$$A = \frac{\text{number of correct transformations}}{\text{total number of test words}} \quad (4)$$

The appropriate cut-off value depends on target word(s) which can be equivalent to the source word. Therefore, it is important that the word generated list of the target is the most probable in the corpus. In this case, a metric that counts the number of translation variants (T_k) that appear in the system-generated list, L might be appropriate.

3.4.2 Multi-Variant Metrics

The corpus can be created using multiple translations, including multiple variants that can be taken into account [2].

Uniform word accuracy (**UWA**) is based on equally values all of the translation variants provided for a source word. For example, consider (S, T) to represent word-pair between source and target, where T = {T_k} and |T| > 1. Therefore, any of the T_k variants in T is successful for translation system.

Majority word accuracy (MWA) is provided as one translation is selected as valid value. The selected valid value as preferred variant it must be suggested by majority of human translators.

Weighted word accuracy (WWA) identifies a weight to each of the translations based on the number of times that they have been suggested with a given weight.

The annotation process can be summarized in terms of the MATTER cycle processes [4]: Model, Annotate, Train, Test, Evaluate and Revise.

3.5. Matter Description

The annotation process can be summarized in terms of the MATTER cycle processes [4]; Model, Annotate, Train, Test, Evaluate and Revise. Figure 2 shows the MATTER development life cycle, [31].

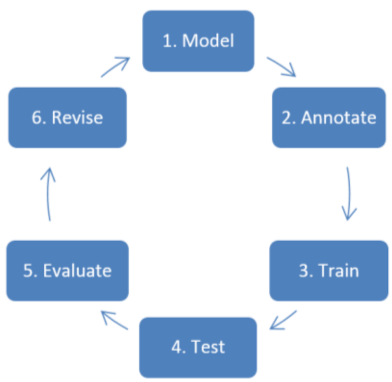

Figure 2: The MATTER Development Life Cycle

The development cycle provides theoretical informed attributes derived from empirical observations over the data. The model can be described by: vocabulary of terms T, the relation between these terms, R, and their interpretation, I. Therefore, the model M can be described by:

$$M = \langle T, R, I \rangle \tag{5}$$

3.6. Generative Translation

Generative translation is the process of translating word or phrase from source language to target language [3]. Many different generative transliteration methods have been proposed in the literature with associated methodologies and languages supported [3]. Automatic transliteration has been studied between English and Arabic [21].

A general diagram of generative translation is shown in Figure 3. Generative-based methods identify the source word S, and then employ the translated evaluation algorithm (single or multi variant) to generate the target word(s) T.

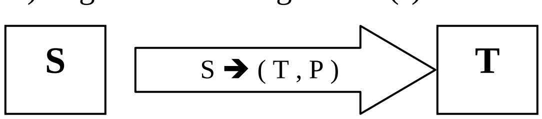

Figure (3): A Graphical Representation Approach

The proposed method of translation system uses an extended Markov window. Such method takes Arabic/English word and uses set of rules then mapped it into English/Arabic target. An alignment method may be used to assign probabilities to set of mapping rules (training stage). The translation model is based on an Markov formula derived from P (S , T) = P(S) P(T|S) as:T = argmax_T P(S) P(T|S)

Choi and et al [19] presented English-Korean transliteration system based on pronunciation and correspondence rules. In such system prefix and postfix was used to separate English words of Greek origin. Also, they designed English-Chinese transliteration frame based model, and used a direct model as explained. Look to the following source-language equation:

$$T = \text{argmax}_T\, P(S|T)\, P(T), \text{ and} \tag{6}$$
$$T = \text{argmax}_T\, P(T|S) \tag{7}$$

They also investigated the target language model to the direct transformation equation as:

$$T = \text{argmax}_T\, (P(T|S)\, P(T) \tag{8}$$

To build their underlying model [3], they presented their model on 46,306 English-Chinese extracted from Linguistic Data Consortium (LDC) entity using word accuracy metrics.
As shown in figure (2), the number of steps in the transformation process is reduced from two or three to one. Such transformation is relying on statistical information using HMM. The following general formula will be used:

$$P(T) = p(t_1) \prod_{i=2}^{m} p(ti \mid ti - 1) \tag{9}$$

Technologies based on NLP are becoming increasingly widespread [18]. Therefore, mobile phones and handled computers support predictive text, lexicon and dictionary building, speech processing and handwriting recognition. Machine translation allows us to retrieve written in language and read them in another language. Consequently, language processing has come to play a central in the multilingual information society. For long time now, machine translation (MT) has been the holy grail of language understanding [5]. Today, practical translation systems exist for specific domains and for particular pairs of languages. According to that natural language toolkit (NLKT) is published and used to support such translation. Many of NLP material are covered in more details [4] [5].

Consequently, simple translator can be made using NLTK by employing source language (e.g. English language) and target language (e.g. French language) pairs, and then convert each to dictionary.

There are many online language translation API's (e.g. provided by Google and Yahoo). Using such API's translation, we can translate text in a source language to a target language. NLTK comes with a simple interface for using it [6]. Therefore, the internet is required to access and used in the translation function. Consequently, to translate text, two things are needed to know:

1. The language of text or source language.
2. The language of want to translate or target language.

4 MOBILE DICTIONARY FRAME WORK

4.1 Principles

The proposed dictionary is a cloud mobile application for an English-English, English-Arabic and Arabic-Arabic dictionaries. The first phase is used to collect and download the data from online English dictionary that is liked “The Project Gutenberg Etext of Webster's Unabridged Dictionary”[10], and it is used to create database file, figure 4.

Figure (4): Dictionary Structure Layout

First the authors classified the dictionary by creating a list of meaning expressions and classifying these meaning in order of their concepts. To classify these expressions the authors need to specify the concepts in the language and define the relations between the words in each concept. The most reasonable classification is the one that are suggested by Hadel and Hassanin [20] [21]. It composed of four main classes: abstracts, entities, events and relations. There are subclasses under each main class and under each subclass may have other subclasses and so on.

Semantic and lexical relations present a suitable way to organize huge amounts of lexical data in ontology's, and other concepts in lexical resources.

4.2 Computing of Mobile Dictionary

It is known that the size of dictionaries database is large and that mobile device storage is small and does not accommodate large amounts of data. The solution for this problem is by using cloud technology. Cloud computing is the use of computing resources such as hardware and software which are existing in a remote location and access such resources and services over a network. The cloud computing service could be divided into three main categories infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS) [25] [26] [27].

There is another category that comes under the three main previous categories, which this paper is interested in; it is data as a service (DaaS). DaaS [28] is a service that makes information and data such as text, image, video and sound reachable for clients through global network. DaaS has many advantages including: reducing overall cost of data delivery and maintenance, data integrity, privacy is satisfied, ease of administration and collaboration, compatibility among diverse platforms and global accessibility. The cloud technology DaaS is used to provide the mobile database for English and Arabic WordNets.

By using cloud technology, the main logical design structure that the mobile dictionary uses will become five tier (layer) structures. The proposed architecture is client/server framework consisting of four layers; each is running on a different platform or in different process space on

[10] http://www.gutenberg.org/cache/epub/673/

the same system. These layers do not have to be physically on different locations on different computers on a network, but could be logically divided in layers of an application [28] [29]. In the four tier structure there are three layers are hidden: presentation layer, process management layer and database management layer. Figure 5 illustrates these four layers. Within the dictionary-scale semantic processing, the cloud computing services; Software as a Service (SaaS), Platform as a Services (PaaS), Infrastructure as a Services (IaaS) [29] and Data as a Services (DaaS) supposed to be employed, as illustrated in figure 5.

The SaaS layer introduces software applications, PaaS presents a host operating system, cloud development tools, while, IaaS delivers virtual machines or processors, supports storage memory or auxiliary space and uses network resources to be introduced to the clients. Finally, DaaS includes large quantity of available data in significant volumes (Peta bytes or more). Such data may have online activities like social media, mobile computing, scientific activities and the collation of language sources (surveys, forms, etc.).

Therefore, cloud clients can access any of the previous web browsers or a thin client with the ability to remotely access any services from the cloud.

4.3 Arabic WordNet Database Design

Arabic WordNet is identical to the Standard English WordNet (PWN and EWN) in structure. Therefore, Arabic words will be organized into four types of POS: nouns, verbs, adjectives and adverbs. Each word is grouped with other words that have the same meaning in a group called Synset. Each Synset is organized under a concept, and it is related to other synset with lexical or semantic relations. Nouns and verbs are arranged in structured way based on the hypernymy/ hyponymy relations. Adjectives are categorized in groups consist of head and satellite synsets. Nearly all head sysnets have one or more synsets that have the same meaning these called satellite synsets. Every adjective is organized based on antonyms pairs. The antonym pairs are in the head synsets of a group.

Figure 5. Proposed Cloud Service Layers.

The proposed database is too big for a mobile device (a mobile application can hold only a database with size 2MB). There are two methods to work with the mobile database, first is locally which is SQLite (offline) and the second uses SQL Azure database (online). The two databases have the same structure but they are different in the data size that they hold. The SQLite database can only hold a small part of the database and can be accessed fast. The SQL Azure database has the whole database and it can be accessed through the internet [30].

4.4 Inter-Lingua in Mobile Dictionary

The proposed system architecture of this paper is based on the interlingua approach in the machine translation (MT). Such approach extracts the meaning of the word from the source language (SL) (English or Arabic) and then translates it in the target language (TL) (English or Arabic). The mechanism can be classified into three main components Arabic language dependent, English language dependent and language independent (inter lingua) modules. Figure 6 explains the proposed mechanism.

The system description includes:

- Bi-lingual dependent modules one for English and the other for Arabic WordNets.
- Domain ontology language independent module to map between Arabic and English WordNets.

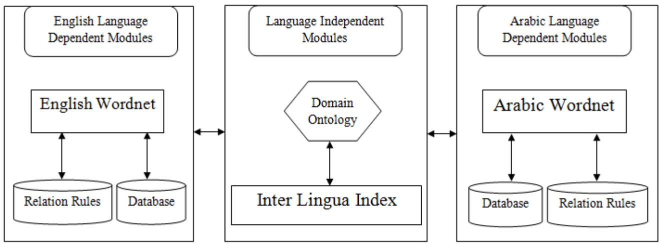

Figure 6. Mobile Dictionary Mechanism

The language dependent modules contain:
1) English language dependent module.

- English WordNet: contains the language vocabularies.
- Lexical Database: this database is described and illustrated in the Princeton WordNet (PWN) [6] [7], which contain approximately most of the English words with their meaning.
- Relation rules: which consist of 16 relations [30].

2) Arabic language dependent module.

- Arabic WordNet: contains the language vocabularies.
- The Arabic lexical database which contains tables that the Arabic database need.
- Arabic Relation rules: include 23 relations types between the synsets: hypernymy, hyponymy, antonym, cause, derived, derived related from, entails, member meronym, part meronym, subset meronym, attribute between adjective and noun, participle, pertainym, see also, similar, troponym, instance holonym, subset holonym, part holonym, instance hyponym, disharmonies, class member and verb group [30].

The language independent module contains:

- Domain ontology: concepts which are grouped in topics by the same. The main goal of the domain ontology is to present a common sense for the most important concepts in all the WordNets.
- The Inter lingua independent (ILI): The goal of the ILI is mapping between the two Synsets of the Arabic and English WordNets.

4.5 Arabic Mobile WordNet (ArWn) Workflow

RESTful web service is used to send and receive data between client and server. The data can be sent and received as Java Script Object Notation (JSON), XML or even as Text. The data of the proposed dictionary is handled by JSON, because it is compact and supported in most of the world.

The RESTful Web services hosted in Windows Azure, it will be used to solve both the interoperability and the scalability in mobile applications. Figure 7 shows the system workflow using RESTful Web service with JSON data format[11]. This workflow is used while taking into consideration hypertext transfer protocol (HTTP), so any client mobile application that supports this protocol is capable to communicate with them; i.e., the interoperability is satisfied. In another direction, windows Azure support scalability to fit any degree of demand of data without difficulty[12].

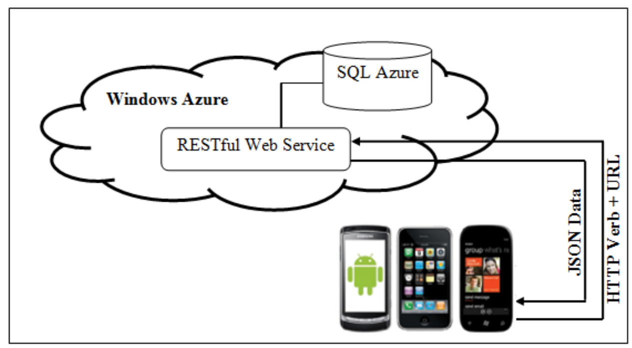

Figure 7. Mobile Dictionary Workflow [30]

3.6 ArWn Implementation Scenario

Implementation steps are divided into four parts:

1. Create an account in the Windows Azure.
2. Build a Windows Azure Cloud Project.
3. Deploy the RESTful Web Service.
4. Build a bilingual mobile application (ArWn).

The WCF REST[13] programming model which is shown in figure 8 permits customization of URIs for all procedures. The model is illustrated in the following:

1. A message request contains an HTTP verb with URL is send from mobile by using standard HTTP.

[11] http://www.slideshare.net/rmaclean/json-and-rest
[12] http://shop.oreilly.com/product/9780596529260.do
[13] http://msdn.microsoft.com/en-us/magazine/dd315413.aspx

2. The RESTful Web service receives the mobile application message request and give a call and pass "$filter=synset_id" as a parameter.
3. Windows SQL Azure database will return the records that are equal to synset_id.
4. The returned data will be converted to JSON format (automatically) and go back to the mobile device.
5. The data will be available to the mobile application.

Three of most widely used mobile operating systems are Apple iOS, Android and Windows Phone. The authors decided to develop the proposed dictionary in an Android platform and Windows phone. Because according to Gartner[14] and IDC[15]. Android is now the most popular and the most used mobile operating system in the world.

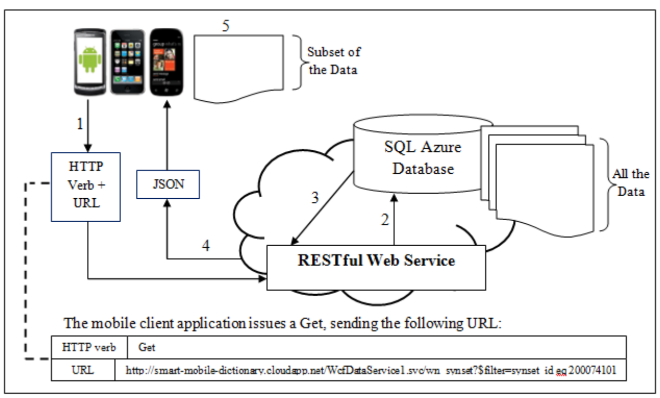

Figure 8. Workflow for Mobile Application Requsting [30]

4.7 Mobile Interface Testing

The authors tried to make the interface of the mobile dictionary user friendly and easy to use as it shown in figure 9. In figure 9, the first screen of mobile is appeared when the application is lunched. The second screen shows all the senses of the word "cat". And the other two screens show an English word "lion" and the equivalent Arabic word "أسد".

5 EVALUATION

The proposed ArWn is made up of Arabic words and related English words, so, the complete synsets includes 5 parts of speech, nouns (6,438), verbs (2,536), adjectives (456), adjective satellite (158), and adverbs (110).

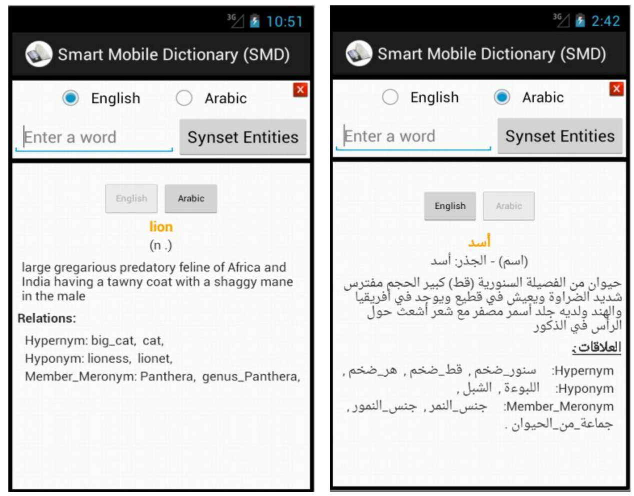

Figure 9. Screens Shoot of the Mobile Dictionary System.

5.1Performance

The response time is important to evaluate the performance of mobile dictionary system. The definition of response time is the duration that a system or application takes to respond to the client. To calculate such time in mobile, we need to know: network bandwidth (speed), number of users (clients), client processing time, server processing time, and network latency time. Therefore, the mobile dictionary system response time can be defined using all the varieties above to return the results to the user (client), as follows:

$$\text{Time} = T_{\text{client}} + T_{\text{network latency}} + T_{\text{server}} \qquad (10)$$

Where:

$$T_{\text{network latency}} = \text{Word meanings} * N / \text{Net Speed} \qquad (11)$$

N represents number of clients.

Real time testing of mobile dictionary is used in order to evaluate the system access time and the needed time to respond and show the results. The testing was done by connecting to the Azure cloud, using Wi-Fi connection with 2MB/S speed.

[14] http://www.gartner.com/newsroom/id/2335616
[15] http://www.idc.com/ getdoc.jsp?containerId= prUS23638712#.USKkKmcV-gM

The service respond time is illustrated in table 1. The respond time is in seconds.

Table 1. Mobile Dictionary Service Respond Time

Service	English word details	Equivalence Arabic details	Arabic word details	Equivalence English details
SQL Azure (Online)				
Word	Rodent	قارض	قارض	Rodent
Time	5.9049 s	5.9066 s	5.1384 s	5.2448 s
Word	Stimulant	منبه	منبه	Stimulant
Time	3.8422 s	3.4754 s	3.1556 s	4.6711 s
Word	Bruise	كدمة	كدمة	Bruise
Time	4.3392 s	3.9299 s	3.0172 s	4.1127 s
Word	Man	رجل	رجل	Man
Time	5.4584 s	4.9957 s	3.7155 s	5.0872 s
Word	Cat	سوط	سوط	Cat
Time	5.1174 s	4.3919 s	3.0083 s	5.7743 s
SQL Azure (Offline)				
Word	Rodent	قارض	قارض	Rodent
Time	0.5215	0.8978 s	0.9635 s	0.5862 s
Word	Stimulant	منبه	منبه	Stimulant
Time	0.3156 s	0.4134 s	0.4084 s	0.3038 s
Word	Bruise	كدمة	كدمة	Bruise
Time	0.3123 s	0.6646 s	0.5485 s	0.3079 s
Word	Man	رجل	رجل	Man
Time	0.7523 s	0.6399 s	0.6333 s	0.6863 s
Word	Cat	سوط	سوط	Cat
Time	0.3130 s	0.5285 s	0.4660 s	0.4818 s

Figures 10 and 11 illustrate the differences between online and offline performance.

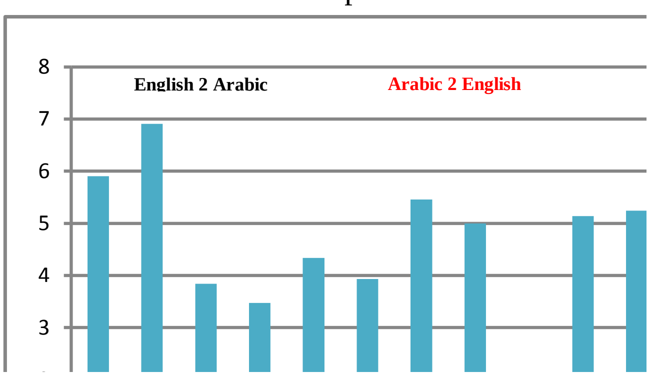

Figure 10. Bi-Lingual Online Translation

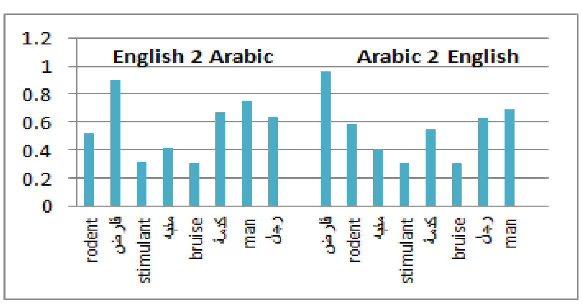

Figure 11. Bi-Lingual Offine Translation

Putting the database in SQL Azure (online) has its advantages and disadvantages. It has been noted from the charts above that extracting the data from SQL Azure takes longer time than extracting it locally from SQLite. Therefore putting the database in cloud database helps to solve the scalability and fixed storage problem in mobile devices but it takes more time to connect to the data.

The proposed ArWn for mobile can be evaluated using semantic relation features. Therefore, this evaluation can be done by linguistic expert. Table 2 illustrates such evaluation results.

Table 2. ArWn Evaluation Features

Semantic Relation	No of Relation	Correct Relation	Precision	Percentage
Attribute	13	11	15.385	84.62
Cause	11	9	18.182	81.82
Class member:Category	10	8	20.000	80.00
Class member:Region	6	4	33.333	66.67
Class member:Usage	3	2	33.333	66.67
Pertainym	12	8	33.333	66.67
Substance holonym	11	8	27.273	72.73
Substance meronym	11	8	27.273	72.73
Member meronym	21	15	28.571	71.43
Member holonym	21	19	9.524	90.48
Part meronym	6	4	33.333	66.67
Part holonym	6	4	33.333	66.67
Hyponym	138	99	28.261	71.74
Instance hyponym	6	4	33.333	66.67
Entails	6	4	33.333	66.67
Antonym	6	5	16.667	83.33
Similar	5	4	20.000	80.00
Derived	26	20	23.077	76.92
See also	6	5	16.667	83.33
Verb group	5	5	0.000	100.00
Participle	3	2	33.333	66.67
Hypernym	37	31	16.216	83.78
Troponym	2	2	0.000	100.00
Disharmonies	2	2	0.000	100.00
Derived related form	2	2	0.000	100.00
Total	375	285	21.35 %	100 %

This evaluation illustrates the value of precession varies from one relation to another due to limited size of dictionary and due to Arabic and English morphological features.

5.2 Cost Evaluation

The proposed mobile dictionary system requires an internet connection to access Azure cloud web server. The Wi-Fi connection is good according to free availability at many locations, especially user's home. The testing proofs that the proposed dictionary system displays good results obtained when testing the application using Wi-Fi connection.

6 CONCLUSIONS

This paper described building bilingual dictionary with lexical and commonsense database. Such dictionary used cloud's technology and services to store the proposed data of the dictionary. Therefore, the authors proposed an application for mobile devices with Android operating system. This application is a dictionary uses the WordNet as a lexical concept and commonsense database. This dictionary is bilingual from English language to Arabic language and vice versa. The RESTful web service of the Windows Azure have been used to deal with the interoperability and data scaling on the storage problem of mobile application.

Moreover, the results of this paper open a new way of approaching for mobile computing in cloud system, by using such technology for reducing the complexity of mobile storage.

In the future, the authors plan to develop the dictionary for other mobile operating system. Also the authors' intent to increase the Arabic language coverage and add to the dictionary some advanced features such as visuality to the Arabic WordNet dictionary. Also, the proposed system can be extended by adding special needs technology, such as sign language, speech recognition and speech synthesis to allow deaf and blind peoples to communicate.

REFERENCES

1. Jurafsky, D. and Martin, J. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall, Englewood Cliffs, NJ, 2008.
2. Karimi, S. Machine transliteration of proper names between English and Persian. Ph.D. dissertation, RMIT University, Melbourne, 2008.
3. Karimi, S. Falk Scholer, F. and Turpin, A. Machine Transliteration Survey. ACM Computing Surveys, Vol. 43, No. 3, Article 17, Publication date: April 2011.
4. Pustejovsky, J., and Stubbs, A., Natural Language Annotation for Machine Learning, 1st Edition, O'Reilly Publisher, Release Date: October 2012.
5. Bird, S., Klein, E. and Loper, E., Natural Language Processing with Python, O'Reilly Media, 2009.
6. Nitin I. and Damerau, F., Handbook of Natural Language Processing, (Second Edition), Chapman and Hall/CRC, 2010.
7. Perkins, J., Python Text Processing with NTK 2.0 Cookbook, Packt Publishing, Birmingham-Mumbai, 2010.
8. Liddy, E.: Natural Language Processing. In Encyclopedia of Library and Inform. Sci. 2nd Ed. Marcel Decker, Inc. 2003, pp. 2126-2136.
9. Hutchins, W.: Machine Translation: A Brief History, Concise history of the language sciences: from the Sumerians to the cognitivists. Edited by E.F.K.Koerner and R.E.Asher, Oxford: Pergamon Press, 1995, pp. 431-445.
10. Tze, L.: Multilingual Lexicons for Machine Translation, ACM, December pp.14–16, 2009.
11. Dichy, J., Farghaly, A.: Roots & patterns vs. stems plus grammar-lexis specifications: on what basis should a multilingual lexical database centered on Arabic be built?, In Proc. of the IXth Machine Translation Summit in the Workshop on Machine Translation for Semitic Languages: Issues and Approaches, New Orleans, USA, Sept. 23, 2003.
12. Weber, G.: Top Languages: The World's 10 Most Influential Languages. Language Today, Vol. 2, Dec. 1997.
13. Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., Fellbaum, C.: Introducing the Arabic WordNet project, In: Proc. of the 3rd Global WordNet Conf., Jeju Island, Korea, 2006, pp. 295-299.
14. Elkateb, S., Black, W., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., Fellbaum, C.: Building a WordNet for Arabic, In: Proc. of The fifth International Conf. on Language Resources and Evaluation; Genoa-Italy, 2006, pp. 29-34.
15. Rodriguez, H., Farwell, D., Farreres, J., Bertran, M., Alkhalifa, M., Martí, M., Black, W., Elkateb, S., Kirk, J., Pease, A., Vossen, P., Fellbaum, C.: Arabic WordNet: Current State and Future Extensions, In: Proc. of the Fourth Global WordNet Conf., Szeged, Hungary. Jan. 22-25, 2008.
16. Rodríguez, H., Farwell, D., Farreres, J., Bertran, M., Alkhalifa, M., Martí, M.: Arabic WordNet: Semi-automatic Extensions using Bayesian Inference, In: Proc. of the 6th Conf. on Language Resources and Evaluation LREC-2008. Marrakech (Morocco), May 2008.
17. Vossen, P.: WordNet, EuroWordNet and Global WordNet, Revue française de linguistique appliquée, Vol. VII, pp. 27-38, 2002.
18. Vossen, P.: Introduction to EuroWordNet, Computer and Humanities, Kluwer Academic Publishers, pp. 32(73-89), 1998.
19. Choi, K.: CoreNet: Chinese-Japanese-Korean WordNet with shared semantic hierarchy, Published in Natural Language Processing and Knowledge Engineering., In: Proc. Int. Conf., Oct. 26-29, pp. 767 – 770, Beijing, China, 2003.
20. Al-Ahmadi, H.: Building ArabicWordNet Semantic-Based Dictionary, master's thesis, Computer Science Dept., King Abdul Aziz Univ., Jeddah, SA, 2010.
21. Al-Barhamtoshy, H., Al-Jideebi, W.: Designing and Implementing Arabic WordNet Semantic-Based, the 9th Conference on Language Engineering, 23-24 December 2009, Cairo, Ain Shams University.

22. Far, R.: Mobile Computing Principles: Designing and Developing Mobile Applications with UML and XML, published by Cambridge Univ. Press, 2005, pp. 861.
23. Talukder, A., and Yavagal, R.: Mobile Computing: Technology, Application & Service Creation, published by the Tata McGraw Hill publishing company limited, Jan 1, 2005, pp. 668.
24. Arokiamary, V.: Mobile Computing, published by Technical Publications Pune, Jan 1, 2009, pp. 556.
25. Strowd, H., Lewis, G.: T-Check in System-of-Systems Technologies: Cloud Computing, Software Engineering Institute, Carnegie Mellon, Pittsburgh, Pennsylvania, Technical Note CMU/SEI-2010-TN-009, 2010, http://www.sei.cmu.edu/library/abstracts/reports/ 10tn009.cfm
26. Lewis, G.: Basics about Cloud Computing, Software Engineering Institute, Carnegie Mellon Univ., 4500 Fifth Avenue Pittsburgh, 2010, http://www.sei.cmu.edu/ library/abstracts/whitepapers/cloudcomputingbasics.cfm
27. Huth, A., Cebula, J.: The Basics of Cloud Computing, Carnegie Mellon Univ., Produced for US-CERT, a government organization, 2011.
28. The ABCs of DaaS- Enabling Data as a Service Application Delivery, Business Intelligence, and Compliance Reporting Revision: 19 September 2011, Delphix Corp.
29. Sadis, F., Mapp, G., Loo, J., Aiash, M., Vinel, A.: On the Investigation of Cloud-based Mobile Media Environments with Service-Populating and QoS-aware Mechanisms, IEEE Transactions on Multimedia, Issue 99, 2013.
30. Al-Barhamtoshy, H., and Mujallid F. Building Mobile Dictionary System, The International Conference on Digital Information Processing, E-Business and Cloud Computing (DIPECC 2013), The society of Digital Information and Wireless Communication (SDIWC), October 23-25, 2013.