Electronic Multilingual Arabic Dictionary Based On Root–Word With Self-Assessment

Muhammad Nasir Ibrahim
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Johor Bharu, Malaysia
mnasir@fke.utm.my

Siti Noormaya Bilmas
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Johor Bahru, Malaysia
snoormaya@yahoo.com

Abuagla Babiker
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Johor Bharu, Malaysia
Abuagla2002@gmail.com

Mariani Idroas
Faculty of Petroleum and Renewable Energy Engineering
Universiti Teknologi Malaysia
Johor Bahru, Malaysia
mariani@petroleum.utm.my

Abstract— This paper describes an electronic multilingual Arabic dictionary that enables searches based on root words or radical alphabets with added self-assessment to enhance learning of the language for non-native speakers. Without the presence of a teacher as available in a school setting, one can develop a language with self-tutoring by using a well designed dictionary with many entries. This is now implemented effectively in an electronic dictionary. Since the majority of Arabic words are trilateral being derived from three radical alphabets, it lends itself naturally to a root-word based organization. However, it is transparent to students who can search full words in any derivative forms. Once a word is found, all necessary information on its meanings, usage, examples, etymology, audio, and other aids can be added in the future. The developed dictionary also includes quizzes which are designed to facilitate students in learning Arabic in effect creating a vocabulary builder.

Keywords- teaching; Arabic; multilingual; learning; self-assessment; root word;

I. INTRODUCTION

There is a need to augment the traditional dictionary even with improved entries with a self-evaluation system such as quizzes centered on the enquired vocabulary. With internet access in education and business in a globalized environment, there is a pressing need for second language acquisition especially so in the Islamic world which unite naturally by the language of the Quran. Arabic is used in twenty Middle East countries as the official language by millions of people. Furthermore, it is also the religious language for all Muslims of various ethnic and cultural backgrounds. Although there exists works on the learning of Arabic by traditional methods in published forms such as dictionaries, linguistic books, and their electronic versions, surprisingly little has been done in the field of computerized language and lexical resources [1].

With the availability of computers to students, and their migration to handheld devices such as smart phones, electronic dictionaries are not only pressing to be made available to every student but to replace the traditional paper formats. This is inevitable because paper is a recurring manufacturing cost, whereas software has zero manufacturing cost with free downloads. Furthermore electronic dictionaries offer facilities which are impossible in paper format such as; interactive student interface, natural language processing, audio cognition, speech recognition and Artificial Intelligence (AI) applications. These capabilities greatly facilitate second language acquisition and they require electronic dictionaries [2].

The traditional arrangement of most paper-based Arabic dictionaries (in the case of Arabic to Malay, all of the known published) is alphabetically, especially those of Arab origins [3]. The other form of arrangement which is common in the Arabic to English dictionaries [4][5] is based on the root word. Although the root-based arrangement is the logical choice for Arabic, otherwise most Arabic words can append an aleph making all word arrangements appearing chaotic and senseless. Unfortunately, this method requires new learners to learn the concept of root words and masdar in order to extract the root word as a prerequisite to search for its meaning, making these kinds of dictionaries inconvenient. Thus, the aim of this paper is to describe a self-assessment Arabic electronic dictionary based on the root word arrangement without needing to learn to extract the root words from compound words murakkab.

In section 2 some theoretical aspects and related work are discussed. Section 3 explores the dictionary design approach and highlights the operational mechanism. Section 4 discusses the issue of Arabic dictionary
implementation. Section 5 explores the methods of assessments while section 6 concludes this paper and highlights future directions related to the issue of the electronic Arabic dictionary.

II. LITERATURE REVIEW

A. Development of Arabic Dictionaries

The existence of many early Islamic works since the codification of the Quran in the eighth century, together with the vast geopolitical and economic influence of the Arabic speaking world naturally created the demand for Arabic dictionaries which are multilingual. Thus, Arabic dictionaries were already in existence before any European ones. Even the basic structure of dictionaries (Muajam and Gamous) originated from the Arabs [6]. Moreover, in Linguistic works, the nation with the most published works is China followed by the Arab countries [6]. However, the development of electronic Arabic dictionaries still lags behind in comparison to its importance, especially free ones which is not in line with the proliferation of smart phones with free Android operating system, and Linux on desktop computers.

B. Problem with Paper based Dictionary

Classical Arabic is a derivative language with almost 100% of words derived from three radicals (trilateral words), and very few being quad literal. Quad literal words are usually loaned words from other languages, for example Injeel which is borrowed from Greek. This statement is easy to verify for classical Arabic because the Quran is the standard Arabic text in which Arabic grammar both classical and modern Arabic adhere to. In fact it is the only language with a standard codex.

In Arabic verb morphology, a root word can be derived to generate many other related words, which represents an expansion of the root. Some expansions are lexical derivations which will result in "new words", and others are variation of the verb's conjugation [7]. Western scholars have assigned Roman numerals to the various patterns of derivation, which are called "forms". The root is designated "Form I".

The main drawback in using the paper root word based dictionary is that the user is required beforehand to extract the root word which may be difficult or impossible for non Arab learners without basic knowledge of Arabic. Therefore, in our implementation, the software will manage the link between compound words murakkab and their roots as well as derivative words. That means the user can search for any word regardless whether it is in root form or otherwise.

Other drawbacks which are associated with non-electronic devices are the absence of interactive interface, and audio-visual aids. On the practicality side, a user often needs to know an answer anywhere and anytime. This is impossible with paper based dictionary as the only device a person carries everywhere today is the Smartphone, which may be the best language tool as it is a spontaneous vocabulary builder, very much similar to how a child learns anything, and especially a mother tongue.

III. RELATED WORKS

Presently, the development of Arabic to Arabic electronic dictionaries has already taken place with state-sponsored programs. The most popular examples are the Interactive Arabic Dictionary on the web (IAD) sponsored by King Abdul-Aziz City for Sciences and Technology (KACST) [8], and the Arab League Educational Cultural and Scientific Organization (ALECSO) (Arab League Educational) with the cooperation with linguists and scholars from Higher Institute for Applied Sciences and Technology (HIAST) [9]. These are successful implementations of interactive Arabic to Arabic dictionary [10] that cater to the needs of Arab learners. Many contributors were involved such as linguistics specialists, scholars as well as computer professionals to produce online versions of the dictionaries.

Albaheth Alarabi [11] represents a rich Arabic to Arabic dictionary. It extracts the meaning of the Arabic word from different famous dictionaries such as Lisanu Alarab, Algamous Almoheet and Alsalah. Although electronic Arabic dictionaries have been around and used by students, a self-assessment for vocabulary improvement has not been developed and integrated. If there exists, it may appear as a small part of an Arabic language exam.

On multilingual electronic Arabic dictionaries, Sakhr [12] implemented a translation system from Arabic to several other languages. Other researchers in this field are making efforts of designing an Electronic Arabic Dictionaries [2][13][14][7]. This shows that a multilingual electronic Arabic dictionary is still an open issue of research and development.

IV. IMPLEMENTATION METHODOLOGY

A. Basic User Enquiry Flowchart

The normal operational mechanism of the dictionary is shown in Figure 1. To access the dictionary, a user needs to open the system either from the computer or online from the dictionary website. Then, the system will ask the user to enter any word. The user needs to make sure that the entered word is valid. If not, the system will remind the user to correct the entry. In future, the system can be improved to provide suggestive words to be chosen. If the word entry is valid, the system will display the meaning of the word and asks the user whether other information related to this word such as derivative words are needed. If so, the system will display the derivative
words with their corresponding meaning, and alternative meanings of the word compounds.

Figure 2 depicts the basic operational flow of the self-based assessment module for improving and enriching the learner’s vocabulary. This module can be integrated with any system for teaching Arabic as a second language. The module is designed to be learner aware in operation so as to adapt to the level of the learners vocabulary with the progressed lessons.

The self-assessment starts with just (multiple choice questions) MCQs, then matches list A with list B, and finally, directs question and answer (Q&A) sessions. The assessment module allows the learner to explore the correct model answers for the wrong ones.

Figure 2 shows three basic parts; first, a login is required for both learner and administrator. Second, the administrator can be required to add exams and monitor the performance of the learner. Third, the student has the freedom to select the type of exams and questions. The learner can then answer, edit, and update the responses.

B. The Internal Structure of the Database Tables

The system is implemented as platform independent; meaning not requiring one particular hardware or software platforms. It is developed using an open source tool; MySQL which is a relational database management system (RDBMS) for the dictionary database and PHP for the web based interface. It can be easily modified, maintained, and upgraded since it is based on open source code.

It consists of two main tables; the first table is the word_meaning table of current collected Arabic words with their corresponding meaning and related information (i.e. root, derivatives and pronunciation), the second table represents the compounded_word_meaning, which contains the explanation of the word in a compounded form. The data dictionary of these tables as well as examples of data content is shown in Tables 1 to Table 4.

A Word_ID is the primary key in the table “word_meaning” and becomes the foreign key in the table “compounded word meaning” in a relationship of
one-to-many. Then, Root_ID is the internal foreign key and it is the unique value of Word_ID that creates the self-joining relationship of one-to-many between the Root_ID and the Word_ID. When the Root_ID is the same as the Word_ID (Root_ID = WORD_ID), this means that the word is a root word. Thus, if someone needs only the root words, it can be achieved by simple query according to the above condition.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Data Type</th>
<th>Length</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word_ID</td>
<td>Number</td>
<td>6</td>
<td>Primary-Key</td>
</tr>
<tr>
<td>Word</td>
<td>Char</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Meaning</td>
<td>Char</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Pron</td>
<td>Char</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Root_ID</td>
<td>Number</td>
<td>6</td>
<td>Refer to word meaning (Word_ID) Self join in (1:N) form</td>
</tr>
</tbody>
</table>

TABLE II. SNAPSHOT OF THE CONTENTS OF (WORD_MEANING) TABLE

<table>
<thead>
<tr>
<th>Word_ID</th>
<th>word</th>
<th>meaning</th>
<th>Pron/ or it can be recorded sound</th>
<th>Root_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>فهم</td>
<td>Understand</td>
<td>الفهم</td>
<td>19</td>
</tr>
<tr>
<td>22</td>
<td>ركب</td>
<td>Ride</td>
<td>ركبة</td>
<td>22</td>
</tr>
<tr>
<td>25</td>
<td>كاتب</td>
<td>Write</td>
<td>كاتب</td>
<td>25</td>
</tr>
<tr>
<td>29</td>
<td>مركب</td>
<td>Boat</td>
<td>مركب</td>
<td>22</td>
</tr>
<tr>
<td>30</td>
<td>كاتب</td>
<td>Clerk</td>
<td>كاتب</td>
<td>25</td>
</tr>
<tr>
<td>32</td>
<td>كتاب</td>
<td>Book</td>
<td>كتاب</td>
<td>25</td>
</tr>
<tr>
<td>33</td>
<td>استفسار</td>
<td>question mark</td>
<td>استفسار</td>
<td>19</td>
</tr>
<tr>
<td>34</td>
<td>فهمي</td>
<td>name (male)</td>
<td>الفهم</td>
<td>19</td>
</tr>
<tr>
<td>46</td>
<td>مكتب</td>
<td>Office</td>
<td>مكتب</td>
<td>25</td>
</tr>
</tbody>
</table>

TABLE III. DATA DICTIONARY OF COMPOUNDED_WORD_MEANING TABLE

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Data Type</th>
<th>Length</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example_ID</td>
<td>Number</td>
<td>6</td>
<td>Primary-Key</td>
</tr>
<tr>
<td>Word_ID</td>
<td>Number</td>
<td>6</td>
<td>Refer to word meaning (Word_ID)</td>
</tr>
<tr>
<td>Example</td>
<td>Var Char</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Example_meaning</td>
<td>Var Char</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

TABLE IV. SNAPSHOT OF THE CONTENTS OF (COMPOUNDED_WORD_MEANING) TABLE

<table>
<thead>
<tr>
<th>Example_ID</th>
<th>Word_ID</th>
<th>Example</th>
<th>Example_meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46</td>
<td>مكتب البريد</td>
<td>post office</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>مكتب السفر</td>
<td>travel agency</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>كتاب الزواج</td>
<td>marriage contract</td>
</tr>
</tbody>
</table>

Figure 3 shows a user enquiring for the meaning of the word كتاب in the ‘Input Interface’. Figure 4 shows the quick output and more options output. Figure 5 shows the detail output interface if the user chooses the second options as shown in Figure 4.

1. You entered the word: كتاب which means: book
2. Click here to get some options such as ‘root’ ‘pronunciation’ and ‘derivation’ ‘example for explaining the meaning.
3. Or Please Click here to search for another word

Welcome, you entered the word: كتاب! which means: book
The family words of the word كتاب starting with the root word are as follows:

- which means........ write ......and pronunciation is....... : kataba Example : .... كتاب كاتب
- which means........ to draw up the marriage contract
- which means........ clerk ........ and pronunciation is....... : katib كتاب
- which means........ book ......... and pronunciation is........ : kitab كتاب الزواج
- which means....... marriage contract
- which means....... writing .......... and pronunciation is.......... : alkitabah كتاب
- which means....... office .......... and pronunciation is.......... : maktabكتب
- which means....... post office
- which means....... library ......... and pronunciation is........ : maktabahكتب
Please click here for more information regarding synonyms and opposite words of كتاب

Figure 6 and 7 show the learner’s tasks and the administrative tasks respectively. Figure 8 shows a snapshot of exam questions (match list_A with list_B and how the student can display the exam and answer it). Figure 9 shows how the student can check the awarded grades and at the same time maximize the learner’s knowledge by displaying the model answer so that the learner can manipulate the wrong answered questions.
The main goal of this system is to accommodate self-tutoring from the self-assessment module on the electronic Arabic dictionary at different learner levels. Therefore, the structure of the user interface to the dictionary is designed not to rely only on the root word as a keyword to search. The system also maintains the relationship between root word and its corresponding derivative words which is an important dictionary structure to enhance the vocabulary of Arabic language students [15]. Thus, this system is grounded on the following basic properties:

1. The structure has been designed to maintain the relationship of the root and derivative words as one family of words so as to develop a sound understanding of their meanings, and usage in their compound words.

2. Enables the user to input in any compound word thus avoids the restriction of relying on the root word as the main keyword for search, which in most parts makes learning Arabic difficult.

3. Enables a self-tutoring assessment module that can help students enrich their vocabulary. Various assessment tools (MCQs, Q&A, List_B with list_B) helps increase the engagement time of the students in understanding the meaning of the words.

4. The translation engine can be adapted to serve other languages and terminologies used in professional fields such as medical, science and technology which are necessary for localization efforts.
V. DISCUSSION

A. Collection

Collection of all possible Arabic lexicons is a monumental task as with any other languages, consisted of classical Arabic and modern Arabic. Classical Arabic are those words collected from manuscripts available from the time of the Quran codex, and those available in the libraries of various Islamic centers of civilization such as in Iraq, Egypt and Spain in the middle ages. Modern Arabic are those words collected from the era of foreign occupation and printed circulations as periodicals after the invention of printers in Europe. Other collections from obscure sources can be added to those collected by Hans Wehr [5], Almujm Alwaseet [15], Lisanu Alarb [16], Alaeen, Al-Mu’jam al-Wasseet [15], Al-Munjid Al-Abjadi [17] and other well known dictionaries. In addition, experts for Multilanguage Arabic dictionaries for selected languages will be engaged for improving, and increasing the content of the translation tables. Cross referencing published works such as Hans Wehr, may be made available after licensing agreements.

B. Arrangement

The main objective of our implementation in this paper is to enable a user to enter a word in any form whether in root masdar or compound murakkab forms, to find the corresponding meaning and other associative information such as its derivatives, and root word.

The structure follows that of MySQL database by maintaining separate tables for links with primary key, foreign key, and other keys that may be added in the future. The table of Arabic lexicon itself is arranged alphabetically based on root word.

VI. CONCLUSION AND FUTURE WORK

Full featured electronic dictionaries help language skills not only for non-native speakers, but also for native speakers to keep them more updated with non-daily used terms. Electronic dictionaries are urgently needed for second language. In the use for second language acquisition, such as Arabic it may replace the traditional paper dictionaries with the penetration of smart phones in which paper format cannot deliver which is interactive interface, audio-visual aids, and self-assessment function. Particularly in Malaysia which established many religious schools some alongside national schools created a natural need for Arabic to Malay dictionaries, and these had existed in large numbers since colonial times. Nevertheless, an electronic Arabic to Malay dictionary as an extension to the Arabic to English dictionary as described in this paper which is platform independent, full featured, open source, and free for all students have not materialized. With many national and international institutions opening up to many Arab students, Arabic has attained a more commercial standing, and more recognized in Malaysia [6]. Therefore, the development of electronic dictionaries as a tool of reference greatly aids in information sharing [7].

In this paper, a framework for an Arabic-to-English dictionary which can be expanded to a multilingual Arabic dictionary, and especially in relevance to Malaysia; Arabic to Malay dictionary has been designed, implemented and tested. Additionally, a self-tutoring module in the form of an efficient set of assessment tools has been designed along with the dictionary to increase the engagement time in encouraging the students to enhance their language vocabulary.

The framework has been designed online on a portable and platform independent tools by using PHP and MySQL. Furthermore, the framework is not restricted to a specific type of dictionary; it can be easily used for several languages using Arabic as an input language. The vocabulary of the dictionary can be updated using a special entry form. Since the implementation is based on table look up, the testing accuracy will be 100% if the requested word exists. The implementation is already multilingual in framework and has hook-up points to expand from Arabic to English to other languages such as Malay, Chinese and Tamil by just changing the output language. Additionally, special purpose dictionaries can be created such as Quran, Medical, Engineering, and other arts and sciences.

To achieve the goal of full featured multilingual Arabic electronic dictionary, future works from this effort can be summarized as follows:

- Integrate this dictionary with the electronic modules for teaching Arabic as second language and design it to adapt according to the user level or according to words in lessons.
- Provide Synonyms and Opposite of the words, and
- Automatic Guessing and suggestion of the required words if the user incorrectly typed the required word. The dictionary can be called from text editors (i.e. Microsoft word, Adobe Readers) to get the meaning whenever the mouse is pointing to the word.
- Use as Arabic spell-checker
- Automatic word pronunciation while reading the text.
- Addition of tashkeel (vowelized) to word search
- Adapt to a smart phones such as Android and Apples iPhone
- Add audio-visual function to enhance learning of words
- Add grammatical comments
- Cross referencing of lexicon with hypertext
REFERENCES


