Automated Tutoring of Arabic Word Root Extraction

Hanan Elazhary, Ahlam Alharthi, Esraa Balkhi, Ghada Aljahdali, Doaa Zagzoog and Atheer Alkhammsh
Computer Science Department, Faculty of Computing & Information Technology
King Abdulaziz University, Jeddah, Saudi Arabia

Abstract—Looking up the meaning of an English word in a dictionary or a lexicon is an easy task. Unfortunately, this is not true in case of the Arabic language in which, typically, the root of a given word has to be derived first before its meaning is looked up. Deriving the root of an Arabic word is not a trivial task. Thus, students always find difficulty in learning the process. Besides, there is hardly any textbook with enough exercises to help students get trained as much as they need. This calls for developing an automated tutor that helps train students in derivating the root of an input Arabic word and generating an explanation of how the root has been derived. This paper presents an automated tool that has been developed for this purpose. It is used to train students to derive the roots of Arabic verbs provided that the root is composed of three or four characters.

Index Terms—Minimum 7 keywords are mandatory. Keywords should closely reflect the topic and should optimally characterize the paper. Use about four key words or phrases in alphabetical order, separated by commas.

1 INTRODUCTION

The Arabic language is characterized by a complex structure. Thus, to look up the meaning of a given Arabic word in a dictionary or a lexicon, the root of the word has to be derived first. This is not a trivial process. Thus, we need an automated tool that can be used as a tutor for helping students get enough training with this process. Many automated tools have been developed in the literature, but most of them aim at generating roots for Natural Language Processing (NLP). This is why each of these tools suffers from one or more of the following disadvantages: (1) does not follow the standard rules of Arabic word root generation, (2) is not error-free, (3) does not generate explanations of how roots are derived and/or (4) cannot be used as a learning tool. Thus, the aim of this research study is to develop an automated tutor that helps train students. In other words, the tool follows the standard rules of Arabic word root extraction, strives to be error-free and generates explanations of how roots are derived to be used as a learning tool. This paper presents the first prototype of the tool that helps derive the roots of Arabic verbs provided that the root is composed of three or four characters. It is accompanied by a lexicon to generate the meaning of a subset of possible verbs (as a proof-of-concept). It is worth noting that another advantage of the proposed tool is that both the explanations and the word meanings can be generated in English so as to be suitable for those learning Arabic as a second language.

The paper is organized as follows: Section 2 presents the related research in the literature. Section 3 provides the details of the Arabic word root extraction rules utilized in the prototype. Section 4 presents the developed prototype and Section 5 presents some additional results (roots of input Arabic words) generated by the prototype. Finally, Section 6 provides the conclusions and directions for future research.

2 RELATED WORK

Al-Shalabi and Evens [1] developed a morphological system for the extraction of the roots of Arabic words. The idea is to remove the longest possible prefix and then to search for a possible root in the remainder by inspecting the possible sets of locations and comparing them to a database of roots. Hawas [2] developed an algorithm that exploits the relations between the word letters and their placement in the word for Arabic root extraction. It tries to predict the word letters comprising the root based on some rules and relations among the word letters and their placement in the word. They claim about 97% success. Yousef et al. [3] developed another algorithm for Arabic word root extraction based on N-gram without using morphological rules in order to avoid the complexity arising from the morphological richness and the multiplicity of morphological rules in the Arabic language. Boudlal et al. [4] utilized a Markovian approach for the extraction of the roots of the Arabic words. Since this is a mathematical approach, it is clear it is not suitable at all as a learning tool. Kanaan and Kanaan [5] developed an algorithm for the extraction of triliteral Arabic roots by examining the words letter by letter starting from the last letter of the word to the first.

These are few of many other proposed systems in the literature. In spite of the diversity of developed systems, they do not achieve 100% accuracy. This has been shown for example by Al-Shawkaf et al. [6] who compared six Arabic root finding algorithms [7], [8], [9], [10], [11], [12]. The reason for this is that most of these algorithms are heuristic algorithms that have been developed for Arabic natural language processing that favors speed over accuracy. Additionally, since they do not follow the exact rules of Arabic word root extraction, they cannot generate corresponding explanations. Thus, these systems are not suitable as learning tools.
3 ARABIC WORD ROOT EXTRACTION RULES

In this prototype, we consider verbs whose roots are of length three and four characters. The rules used for the extraction of such roots are explained below.

3.1 Rule 1
Rule 1 states that if the verb is composed of four characters and the second and third characters are repeated, then we delete one of the duplicate characters. For example the root of عمل is تعلُّم. It is worth noting that the ◦ indicates duplication of the corresponding character.

3.2 Rule 2
Rule 2 states that if the verb is composed of three characters and the first character is ١, then we replace ١ and ١. For example the root of أكل is أكل أفْرَقٌ. And the second character is ١, then we replace ١. For example the root of جُلُد is جُلُد. And the third character is ١, then we delete ١. For example the root of جُلَّد is جُلَّد. And the fourth character is ١, then we delete ١. For example the root of أَجُلّد is أَجُلّد. And the last character is duplicated, then we delete the first character is ١, then we delete ١, and the second character is ١, then we delete ١. For example the root of جُلَّد is جُلَّد.

3.3 Rule 3
Rule 3 states that if the verb is composed of six characters, the first character is ١, and the fourth character is duplicated, then we delete ١ and the duplicate characters. For example the root of أَطْمَأْنَ is أَطْمَأْنَ.

3.4 Rule 4
Rule 4 states that if the verb is composed of six characters, the first character is ١, and the fourth character is ١, then we delete ١, and the last character is ١. For example the root of أَطْمَأْنَ is أَطْمَأْنَ.

3.5 Rule 5
Rule 5 states that if the verb is composed of six characters, the first character is ١, the fourth character is ١, and the last character is duplicated, then we delete the first ١ character, the fourth ١ character and one of duplicate characters. For example the root of أَطْمَأْنَ is أَطْمَأْنَ.

3.6 Rule 6
Rule 6 states that if the verb is composed of six characters, the first character is ١, and the last character is duplicated, then we delete the ١ and one of duplicate characters. For example the root of أَطْمَأْنَ is أَطْمَأْنَ.

3.7 Rule 7
Rule 7 states that if the verb is composed of six characters, the first character is ١, the fourth character is ١, and the third and fifth characters are similar, then we delete ١, ١, ١, ١, and one of the similar characters. For example the root of عُشَبٌ is أَعْوَدْشَبٌ.

3.8 Rule 8
Rule 8 states that if the verb is composed of five characters, the first character is ١, and the fourth character is duplicated, then we delete the ١ and one of the duplicate characters. For example the root of حُمَرٌ is أَحْمَرٌ.

3.9 Rule 9
Rule 9 states that if the verb is composed of six characters, the first character is ١, the second character is ١, and third character is ١, then we delete the ١, ١, ١, ١, and one of the duplicate characters. For example the root of خَرَجٌ is أَخْرَجٌ.

3.10 Rule 10
Rule 10 states that if the verb is composed of five characters, the first character is ١, and the third character is ١, then we delete ١ and ١. For example the root of عَمَلٌ is أَعْمَلٌ.

3.11 Rule 11
Rule 11 states that if the verb is composed of five characters, the first character is ١, and the third character is duplicated, then we delete ١ and one of the duplicate characters. For example the root of ١ is أَثْمٌ.

3.12 Rule 12
Rule 12 states that if the verb is composed of five characters and the first character is ١, then we delete ١. For example the root of ١ is أَحْرَجٌ.

3.13 Rule 13
Rule 13 states that if the verb is composed of five characters, the first character is ١, and the second character is ١, then we delete ١ and ١. For example the root of أَكْمِر is أَكْمِر.

3.14 Rule 14
Rule 14 states that if the verb is composed of four characters and the second character is ١, then we delete ١. For example the root of أَقْلٌ is أَقْلٌ.

3.15 Rule 15
Rule 15 states that if the verb is composed of five characters, the first character is ١ and the third character is ١, then we delete ١ and ١. For example the root of تُنْصَرٌ is أَتْنَصَرٌ.

4 THE PROTOTYPE

The developed prototype is an automated tutor that accepts any of the verbs whose root is of size three or four and generates the root. It generates an explanation of how the root has
been extracted so that students can benefit from the explanation. Explanation can be generated in both Arabic and English so that Arabic students and students who are learning Arabic as a second language can utilize it. The prototype is accompanied by a dictionary that can also generate the meaning of the input word depending on the generated root. Both formal and simplified explanations are generated. Figures 1 through 6 show snapshots of the system.

5 Results

This section provides results of the prototype. In other words, it provides additional examples for some roots that have been derived using the prototype based on the utilized rules. The results are shown in Table I.
6 CONCLUSION

This paper proposed developing an automated tutor that can be used for training students to extract the roots of Arabic words. The tutor accepts an input word and generates the corresponding root, explanation of how the root has been extracted and the meaning of the word. The tool generates both the formal and simplified explanations. The tool has an English interface in addition to the Arabic one to be usable by students learning Arabic as a second language.

So far, a prototype has been developed. It considers Arabic verbs whose roots are of length three or four. In the future, the prototype will be extended to cover all root generation rules and will be accompanied by a comprehensive lexicon. The tool will also be tested empirically and the results will be presented in subsequent papers.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Input Word</th>
<th>Extracted Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule 1</td>
<td>كتاب</td>
<td>كتاب</td>
</tr>
<tr>
<td>Rule 2</td>
<td>أخذ</td>
<td>أخذ</td>
</tr>
<tr>
<td>Rule 3</td>
<td>عظوز</td>
<td>عظوز</td>
</tr>
<tr>
<td>Rule 4</td>
<td>جرحهم</td>
<td>جرحهم</td>
</tr>
<tr>
<td>Rule 5</td>
<td>اختصار</td>
<td>اختصار</td>
</tr>
<tr>
<td>Rule 6</td>
<td>نعشر</td>
<td>نعشر</td>
</tr>
<tr>
<td>Rule 7</td>
<td>اختشن</td>
<td>اختشن</td>
</tr>
<tr>
<td>Rule 8</td>
<td>مصر</td>
<td>مصر</td>
</tr>
<tr>
<td>Rule 9</td>
<td>استغر</td>
<td>استغر</td>
</tr>
<tr>
<td>Rule 10</td>
<td>قبل</td>
<td>قبل</td>
</tr>
<tr>
<td>Rule 11</td>
<td>قدم</td>
<td>قدم</td>
</tr>
<tr>
<td>Rule 12</td>
<td>تجزيب</td>
<td>تجزيب</td>
</tr>
<tr>
<td>Rule 13</td>
<td>اجل</td>
<td>اجل</td>
</tr>
<tr>
<td>Rule 14</td>
<td>فعل</td>
<td>فعل</td>
</tr>
<tr>
<td>Rule 15</td>
<td>جمع</td>
<td>جمع</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENT

The authors would like to thank Ahmed Alsherbin, Bachelor of Arabic Language, Suez Canal University for his help regarding the Arabic root extraction rules utilized in the proposed automated tutor.

REFERENCES


