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 by deriving 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

THIS 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 handles all these issues. 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 meaining 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-Shawakfa 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.

_____

• *Hanan Elazhary is an associate professor in the Computer Science Department, Faculty of Computing & Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. E-mail: helazhary@kau.edu.sa*
• *The rest of the authors are undergraduate students in the same department*

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 علّم is علم. 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 آ, then we replace آ with أ. For example the root of آكل is أكل.

3.3 Rule 3

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

3.4 Rule 4

Rule 4 states that if the verb is composed of six characters, the first character is ا and the fourth character is ن, then we delete ا and ن. For example the root of افرنقع is فرقع.

3.5 Rule 5

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

3.6 Rule 6

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

3.7 Rule 7

Rule 7 states that if the verb is composed of six characters, the first character is ا, the fourth character is و and the third and fifth characters are similar, then we delete ا, و and one of the similar characters. For example the root of اعشوشب is عشب.

3.8 Rule 8

Rule 8 states that if the verb is composed of five characters, the first character is ا, and the fourth character is duplicated, then we delete the ا and one of the duplicate characters . For example the root of احمرّ is حمر.

3.9 Rule 9

Rule 9 states that if the verb is composed of six characters, the first character is ا, the second character is س and third character is ت, then we delete the ا, س and ت. For example the root of استخرج is خرج.

3.10 Rule 10

Rule 10 states that if the verb is composed of five characters, the first character is ت and the third character is ا, then we delete ت and ا . For example the root of تعامل is عمل.

3.11 Rule 11

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

3.12 Rule 12

Rule 12 states that if the verb is composed of five characters and the first character is ت, then we delete ت. For example the root of تدحرج is دحرج.

3.13 Rule 13

Rule 13 states that if the verb is composed of five characters, the first character is ا and the second character is ن, then we delete ا and ن. For example the root of انكسر is كسر.

3.14 Rule 14

Rule 14 states that if the verb is composed of four characters and the second character is ا, then we delete ا. For example the root of قاتل is قتل.

3.15 Rule 15

Rule 15 states that if the verb is composed of five characters, the first character is ا and the third character is ت, then we delete ا and ت. For example the root of انتصر is نصر.
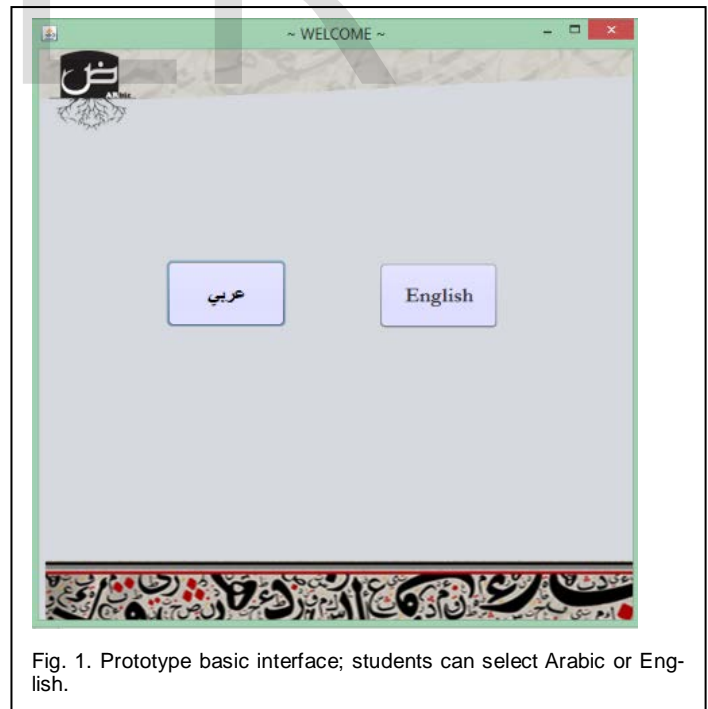

Fig. 1. Prototype basic interface; students can select Arabic or English.

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.
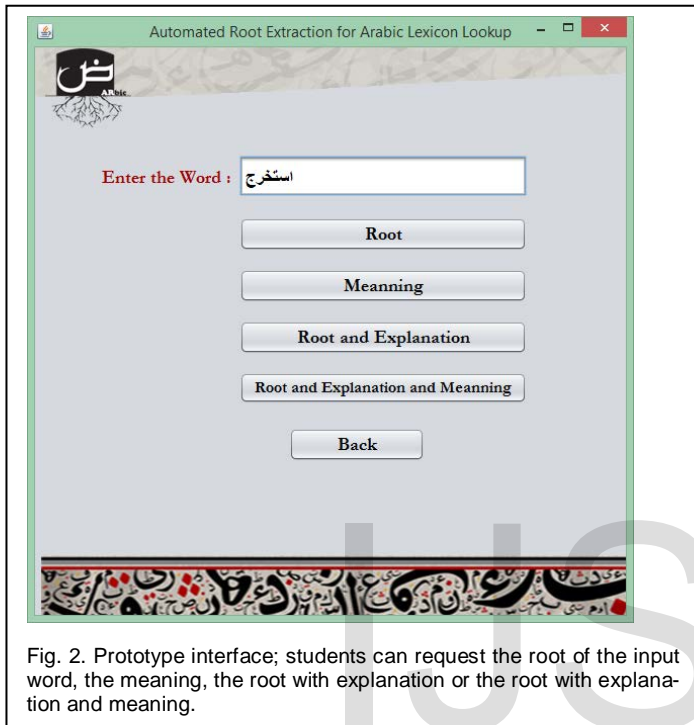

Fig. 2. Prototype interface; students can request the root of the input word, the meaning, the root with explanation or the root with explanation and meaning.
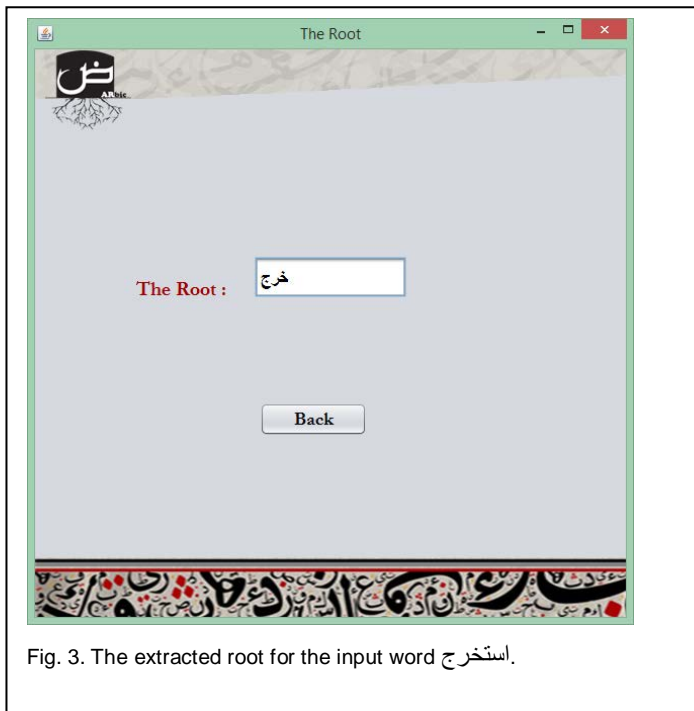

Fig. 3. The extracted root for the input word استخرج.
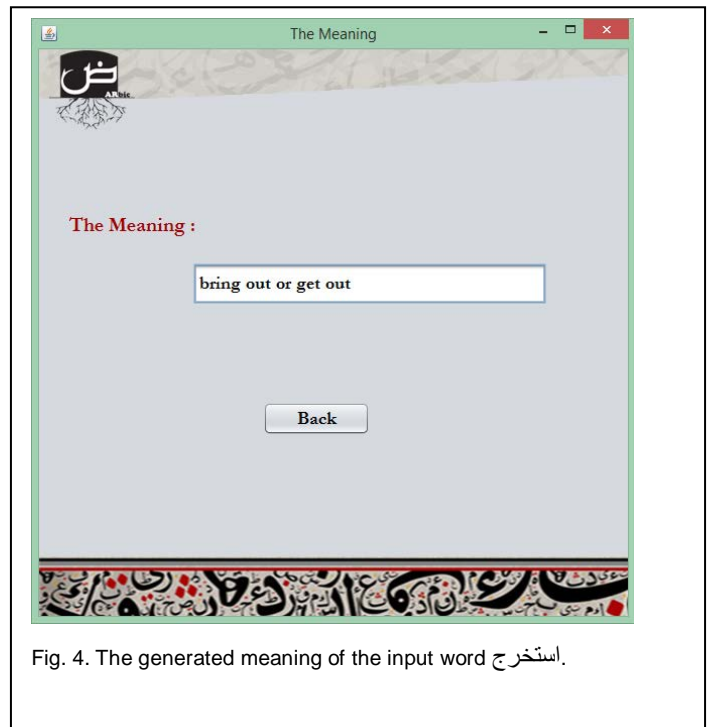

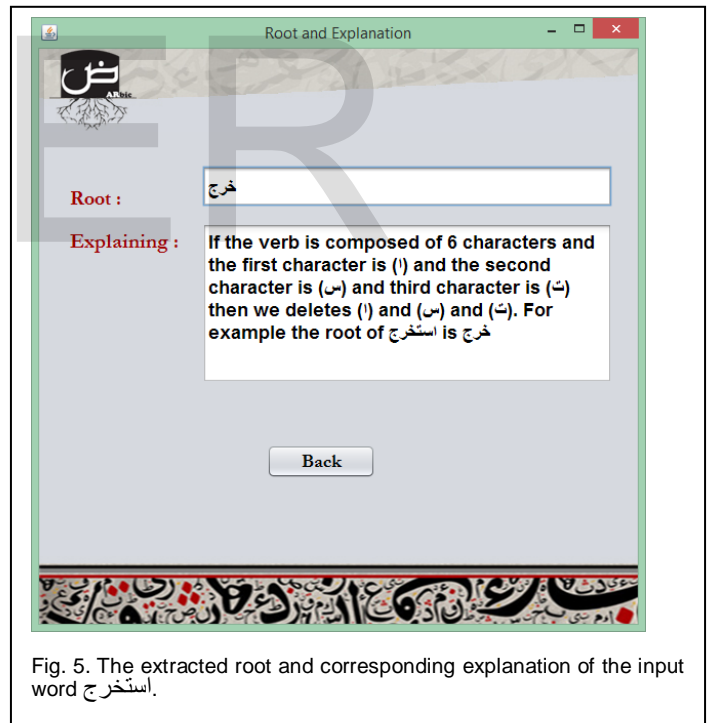Fig. 4. The generated meaning of the input word استخرج.

Fig. 5. The extracted root and corresponding explanation of the input word استخرج.

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.
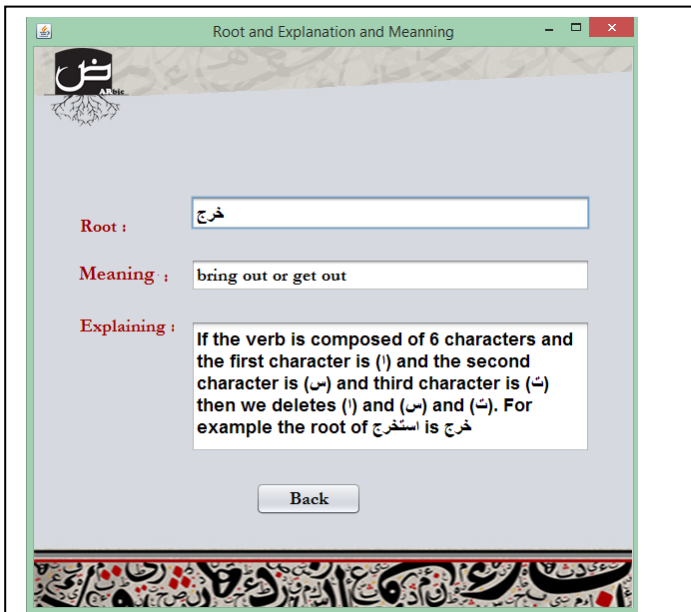
Fig. 6. The extracted root and corresponding meaning explanation of the input word استخرج.
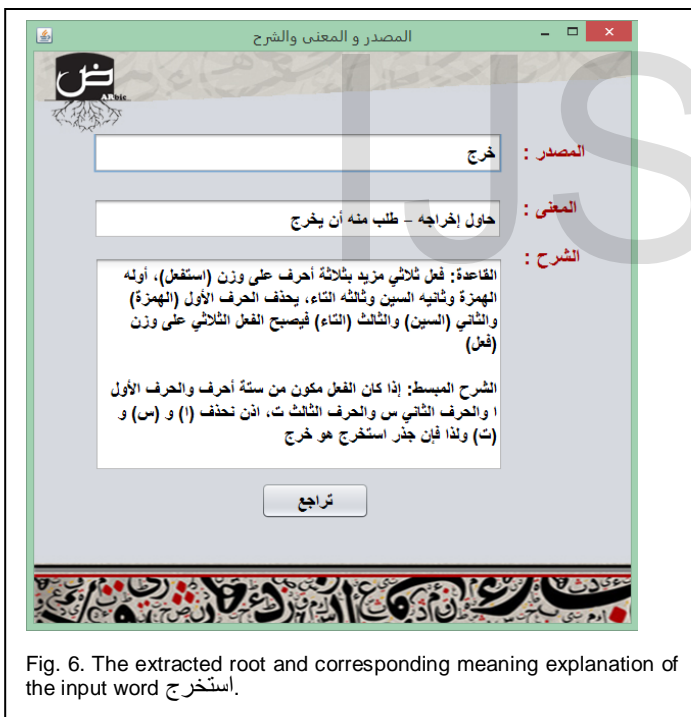

Fig. 6. The extracted root and corresponding meaning explanation of the input word استخرج.

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 simplifed 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 1
RESULTS OF THE PROTOTYPE

Rule	Input Word	Extracted Root
Rule 1	كذَّب	كذب
Rule 2	آخذ	أخذ
Rule 3	اعلوّط	علط
Rule 4	احرَنْجَم	حَرجَم
Rule 5	اخضارّ	خضر
Rule 6	اشمأزّ	شمأز
Rule 7	اخشوشن	خشن
Rule 8	اصفرّ	صفر
Rule 9	استغفر	غفر
Rule 10	تقابل	قبل
Rule 11	تقدّم	قدم
Rule 12	تجوْرَب	جوْرَب
Rule 13	انغلق	غلق
Rule 14	فاعل	فعل
Rule 15	اجتَمع	جمع

REFERENCES

[1] R. Al-Shalabi and M. Evens, "A Computational Morphology System for Arabic," *Proc. the Workshop on Computational Approaches to Semitic Languages*, pp. 66-72, 1998.

[2] F. Hawas, "Exploit Relations between the Word Letters and their Placement in the Word for Arabic Root Extraction," *Computer Science*, vol. 14, no.2, pp. 327-341, 2013.

[3] N. Yousef, A. Abu-Errub, A. Odeh, and H. Khafajeh, "An Improved Arabic Word's Roots Extraction Method Using N-gram Technique," *Journal of Computer Science*, vol. 10, no. 4, pp. 716-719, 2014.

[4] A. Boudlal, R. Belahbib, A. Lakhouaja, A. Mazroui, A. Meziane, and M. Bebah, "A Markovian Approach for Arabic Root Extraction," *The International Arab Journal of Information Technology*, vol. 8, no. 1, 2011.

[5] R. Kanaan and G. Kanaan, "An Improved Algorithm for the Extraction of triliteral Arabic Roots," *European Scientific Journal*, vol. 10, no. 3, 2014.

[6] E. Al-Shawakfa, A. Al-Badarneh, S. Shatnawi, K. Al-Rabab'ah, and B. Bani-Ismail, "A Comparison Study of Some Arabic Root Finding Algorithms," *Journal of the American Society for Information Science and Technology,* vol. 61, no. 5, pp. 1015-1024, 2010.

[7] S. Khoja and R. Garside, "Stemming Arabic Text," *Lancaster University*, Computing Department, 1999.

[8]   R. Al-Shalabi, G. Kanaan, and H. Al-Serhan, "New Approach for Extracting Arabic Roots," Proc. International Arab Conference on Information Technology, pp. 42-59, 2003.

[9]   K. Taghva, R. Elkhouri, and J. Coombs, "Arabic Stemming without a Root Dictionary," *Proc. the International Conference on Information Technology: Coding and Computing*, pp.152–157, 2005.

[10] R. Al-Shalabi, "Pattern-Based Stemmer for Finding Arabic Roots," *Information Technology Journal*, vol. 4. no. 1, pp. 38-43, 2005.

[11] S. Ghwanmeh, R. Al-Shalabi, G. Kanaan, K. Khanfar, and S. Rabab'ah, "An Algorithm for Extracting the Root for the Arabic Language," *Proc. the Fifth International Business Information Management Association Conference*, 2005.

[12] R. Sonbol, N. Ghneim, and M. Desouki, "Arabic Morphological Analysis: A New Approach," *Proc. the Third International Conference on Information and Communication Technologies: From Theory to Applications*, pp. 1–6, 2008.

IJSER