IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 2, 2019, accepted February 25, 2019, date of current version April 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2908177

Root Identification Tool for Arabic Verbs

BAKEEL AZMAN
College of Computer and Information Technology, Sana'a University, Sana'a, Yemen
e-mail: bakeel.azman@su.edu.ye

ABSTRACT Numerous Arabic morphology systems have been devoted toward morphed requirements of words that are required by other text analyzers. Term rooting is an essential requirement in those systems, yet rooting module in the state-of-the-art morphology systems insufficiently meet that requirement, especially verb term. Consequently, due to termination in stemming term rather than a rooting term. Since the stem of the verb is not the root of the verb, it is not feasible to generate or inference verb's derivations and whole it's surface forms (patterns) such tense, number, mood, person, aspect, and others of verb irregular patterns. Therefore, we propose a new model for identifying the verb's root produced in a tool (RootIT) in order to overcome verb root extraction without disambiguation out of traditional methods, applied in current morphology systems. A major design goal of this system is that it can be used as a standalone tool and can be integrated, in a good manner, with other linguistic analyzers. The adopted approach is a mapping surface verb with full-scale derivative verbs discharged previously in the relational database. Moreover, the proposed system is tested on the adopted dataset from PATB verbs extracted from CoreNLP system. The extracted dataset, containing more than (7950) distinguishes verbs belonging to (1938) different roots. The results obtained outstrip the best-compared system by (2.74%) of high accuracy.

INDEX TERMS Root, verb, pattern, stem, morphology, identifying, and ANLP.

I. INTRODUCTION

The Arabic sentence falls under one of two types, either nominal or verbal sentence [1]. Their existence in the text can be of nearly equal proportion, on the understanding that it is possible to convert one type to another. In respect to verbal sentence, it cannot be formed without having a verb which classifies that sentence as a verbal sentence. The term of verb in that type of sentence is our main concern in this study. Arabic and other Semitic languages have long been described in terms of a root interwoven with a pattern. The root is a sequence of consonants, as each Arabic verb contains (3) or (4) consonants that generally remain unchanged in all of its conjugated forms and form the consonantal root; all the remaining information on a conjugated form is called "pattern" such as (مفعل، يفعل، فعل إلخ...،تفاعيل) [2]. The patterns morph the meaning of the root to create a variety of related words. Thus, the task of rooting verb constitutes a cornerstone for all morphology systems and almost linguistic analyzers systems that cares about events (verbs) along with its variant surface forms. Furthermore, systems aspire to obtain a verb conjugation, other systems require a verb's surface form such parsers, and other desire to provide alternative verb's forms in order to paraphrase a fragment text along without violating the meaning. In addition, task of semantic analysis needs to collect sentence arguments on the basis of thematic their role, which has more importance as a system [3]. The point, which motivated this study, is our own work that intends to parse the cognate object that occurs in Arabic sentence, as we have been in serious need to get the verb's root of which spans the cognate clause.

Rooting studies in a specialized manner enrich basically the morphology systems with one of the most important morphed feature. In addition to supplying analyzers with a special tool concerning term roots along with all their patterns and surface forms. Embedding such a tool within related applied systems play pivot roles in the manipulation of sentence elements, out of these systems are information extraction, in Ontology relations selection (predicate) [4], in lexicon and dictionary, in machine translation [5], [6], in subject-verb agreement and object-verb agreement, in verb movement, in verb sense disambiguation, Arabic spell checking, diacritizing (تشكيل) and so forth.

The extraction of verbs roots is important not only for a pure linguistic analysis, but also for discovering different arguments of sentence, purpose, and effect in speech, and for more-sophisticated Natural Language Understanding systems in general. In spite of this importance, there is no

The associate editor coordinating the review of this manuscript and approving it for publication was Mohamed Elhoseny.

specific tool or sufficient embedded module stand up to identify word's root.

Our proposed model is specialized in the extraction of real verb root existing in its original (3) or (4) characters distancing busy with other morphed features, although it is possible to include those features due to modeling of root and its surface forms in this model. The main method of RootIT tool depends on tree-structure hierarchized for each verb root, starting by root as tree root, following some levels ending with leave that present real verb form in real text. The whole surface forms of the verbs have been derivated via ALECSO (Arab League Educational, & Scientific and Cultural Organization) derivational system Sarf [7]. By this method, all derivated and all surface forms of verb roots are represented in relational database, in this way, the ability of extracting any verb root will be fully performed. The work area will be performed on any text and on any form of verb without constraints on inputs as happening in similar systems.

A. ROOT vs STEM

A root is a form, which is not further analyzable, either in terms of derivational or inflectional morphology [8]. Additionally, it is that part of word-form that remains when all inflectional and derivational affixes are removed. A root is also the basic part, which is always present in a lexeme. While the stem is a root of a word, together with any derivational and inflectional affixes are added [9]. A stem consists minimally of a root, but may be analyzable into a root plus derivational morphemes. A stem may require an inflectional operation (often involving a prefix or suffix) in order to ground it into discourse and make it a fully understandable word. If a stem does not stand by itself in a meaningful way in the language, it is referred to as a bound morpheme.

Furthermore, there is a term called ''base'' which is any form to which affixes of any kind can be added [10]. This means that any root or any stem can be termed as a base, but the set of bases is not exhausted by the union of the set of roots and the set of stems: a derivationally analyzable form to which derivational affixes are added can be only referred to as a base. That is, يستهدف ''target'' can act as a base for a prefix to give سيستهدف ''will target'', but in this process يستهدف could not be referred to as a root because it is analyzable in terms of derivational morphology, nor as a stem since it is not the adding of inflectional affixes which is in question [11].

A root differs partially from a stem in that a stem must have lexical meaning. A root has no lexical meaning and the semantic range of the root is vague if there is any. A stem may contain derivational affixes and it becomes of concern only when dealing with inflectional morphology. In the form سيستهدفون ''They will target'' the stem is ''يستهدف'', although in the form ''يستهدف'' the stem is ''يهدف''. Stemming sometimes affects the semantic of a word, whereas lemma preserve the meaning of a word [12].

Root, stem and base are all terms used in the literature to designate that part of a word that remains when all affixes are removed. The distinction is only useful in a highly inflected language like Arabic. In English, both words are used in the same way, to indicate which one can have an affix. Since there are very few affixes in English, it doesn't really matter.

II. RELATED WORK

Current morphology systems seek to get the stem of word rather than the. They endeavor for obtaining the stem of a term as it looks like a non-affixed term such MADAMIRA [13] that adopts some attached toolkit producing stem, based on decomposition of the input word e.g. suffixes and prefixes along with clitics. The extraction rules have failed in getting sound root e.g. ٱِسْتَنْكَر ''denounce'', اِرْتَكَبَها ''commit'', تَمَنِّي ''hope''and تَلَطَّخ ''be stained''. All morphology systems are supposed to make a difference between stem and root [14]. The potential root of a term should be casted into its original orthography of letters, just trilateral or quadrilateral atom.

The list of Arabic stemming proposed models is huge [15]–[22]; thus, it is impossible to review them briefly in this section. Notwithstanding, we will review the main approaches adopted in common models and highlight state-of-the-art in the same subject. The dominant approaches in extracting roots systems are two, light stemmer and root-based stemmer [23]. Stem-based algorithms remove prefixes and suffixes from Arabic words, while root-based algorithms reduce stems to roots [24]. Light stemming refers to the process of stripping off a small set of prefixes and/or suffixes without trying to deal with infixes or recognize patterns and find roots [14], by very briefly word, Light stemmer cannot deepened for extracting of root [25], [26]. The second approach is root-based stemmers, whereas the name implies that root is extracted from the word by means of morphological analysis. It attempts to restore original root of a word and group words accordingly [23]. The basic two steps of root-based stemmers are first to remove prefixes, and suffixes. Second, is to extract roots by analyzing Arabic words according to their morphological components. This is accomplished by rule based techniques, table lookup [25], or by a mixture of the two.

One of the earliest techniques developed for root-based stemmers is Khoja stemmer [20]. In this technique, prefixes and suffixes are removed, then two dictionaries are used, one to match the remaining letters against Arabic patterns, and the second is to confirm the correctness of the root. Taghva *et al.* [27] is similar to Khoja *et al.* with no use of dictionaries; rather a rule-based technique is used. Sonbol *et al.* [28] uses a rule-based technique to extract roots by dividing letters to a part of root, and others are further divided into sub-groups which are examined with well-defined rules to extract the final root, with no use of a dictionary. Moreover, Spline function technique is utilized for extracting root. That method is divided into two phases. First phase involves seeking all the possible roots of each term analyzed out of the context with a morphalizer. Second phase constructs a disambiguation approach based on continuous quadratic splines to choose among these roots the one that corresponds to the word context [29]. Momani and Faraj [30] filters rootless words, and then removes suffixes and prefixes. It removes excessive letters سألتمونيها only if it takes place more than once in a word.

A new stemming model is introduced to design and implement an Arabic light stemmer which claimed that it can identify the word root. It uses predefined mathematical rules and several relations between letters of the term. After applying an appropriate rule on a word, some clitics will be removed, then mapping the produced word against roots dictionary. If no mapped root in the dictionary, splitting the original word process repeats through another rule. The process continues recursively until finding a root or last category rules with no root [31]. However, the rules set cannot face multitude of Arabic word morphology, as evidenced by some unsound words (يستلهمون, تتوقد,يستخدمون, ..., etc.). Mohammed [32] conceived the two approaches problems, so he proposed an Arabic stemmer that combined the rules of root-based stemmer and light-based stemmer to success in facing up that failures. Such problems are removing the affix before matching it with a *Tafeala* تفعيل and not dealing with the word of three letters length, resulted from these two stemmers. A model attempted to mix the two approaches along with statistical approach by [33]. It was only for generating the possible roots of any given Arabic word. The analyzer is based on automatically derived rules and statistics. It compounded three modules, one for taking advantage of a list of Arabic word-root pairs to derive a list of prefixes and suffixes, another one for building stem templates, and last one for calculating the possibility that a prefix, a suffix, or a template would appear. The second accepts Arabic words as input, attempting to construct possible prefix-suffix temple combinations, and outputs a ranked list of possible roots.

The most relevant method is [34], which attempts to build a lexicon including all verbal forms. The verbal forms are generated from more than (15000) roots which is an undefined source. The generation process is applied based on root-patterns using finite-state transducers theory. Almost 2.5 million verb forms are generated and classified within a lexicon. Nevertheless, the verification of produced forms remains unclear, thus the validation of forms cannot be judged. That is because process of test is implemented on small fragment of Nemlar [35] corpus which itself is not more than (500000) words, including nouns and adjectives according to his claims. Many systems, related to this research, have been studied, but unfortunately we have not come across a model that includes all the basic and standard roots in the thesaurus, such as the Al-Mukhtar Al-Sahah[1] dictionary, which contains (7400) and more of roots, although the used ones do not exceed (3000) roots.

In general, extracting root systems are mostly relying on rule-based approach [36], which in turn has deficiencies in tackling abundance of Arabic language inflections and derivations. Since most of the rule-based Arabic stemmers go to remove prefixes, infixes, and suffixes, most of these stemmers cannot not recognize the right root of some Arabic terms, i.e. يستهدفون ''they target'', تتحلل and ''decompose'', استنطقها ''interrogated''. Majority drops takes place in vowel-words with one of ا و ي letters. All those Arabic words do not include all the letters of the corresponding Arabic root, so rule-based Arabic stemmers are incapable to extract their correct roots.

Up to now, there is no available appropriate tool for Arabic NLP applications fully meets requirements of word root. Although many research projects have focused on the problem of Arabic morphological analysis using different techniques and approaches as we most briefly discussed fragment models.

III. PROPOSED WORK

The system of RootIT for verb root identification is a root-based system, and it contains two modules, roots database module and module of detecting of root. The general approach in most morphology systems are top-of-bottom, starting of capturing a sentence word w, then mapping w against embedded lexicon, if no mapped, the cycle evokes trimer of preffixations or suffixation. The process continues recursively, trimming, mapping, until w matches a lexicon entry which present the stem of w. Our approach seems tiny different; it adopts bottom-up approach along with top-of-bottom. The method centerizes the root r in the center of tree, then the root r is reproduced its conjugations and so surface forms surrounded with levels. The leave level (real verb) indicates to its root center via tracing path which will present the root of the verb v. FIGURE 1 shows method in an example for هدف ''aim'' root.

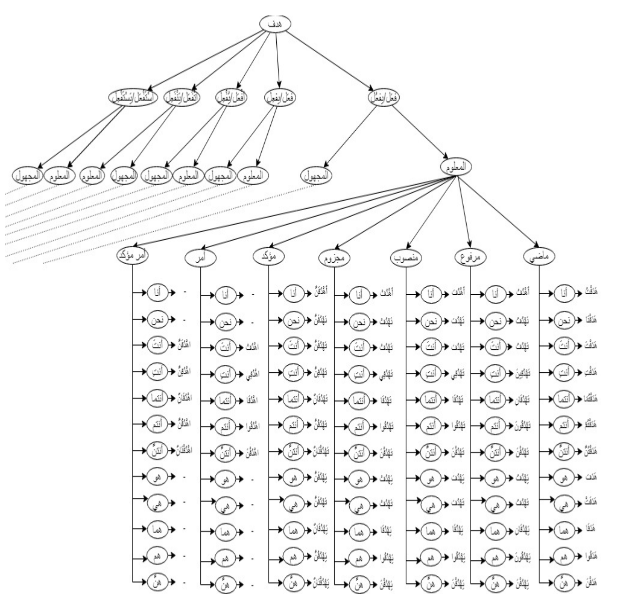

FIGURE 1. Tree-hierarchized structure for هدف. "aim" root.

[1] https://www.almaany.com/ar/dict/ar-ar/?c=%D9%85%D8%AE%D8%AA%D8%A7%D8%B1%20%D8%A7%D9%84%D8%B5%D8%AD%D8%A7%D8%AD c

Detecting root is the function of surface form verb and its path. Given a verb *v*from text fragment, we can retrieve matched verbs against roots database. Since *v*is mostly non-vocalized word, the result of query of search *v*retrieves list of matched entries, that have variant roots. The matched verbs are clustered in clusters of roots, as the cluster head is the root while cluster children are the matched verbs. The root of *v* is the cluster that has maximum number of children. As an example, consider verb تعلن "it announces" retrieved more than one root (علي, علل, عيل, عول,علن). The sound one is علن "announce", so large retrieved in the list of verbs (children of clusters) is for علن and less for rest clusters.

Development of RootIT tool seems straightforward without complexity. RootIT's method to identify verb root, that adopts mapping method the real verb in text against surface verb forms stored in database as shown in tool architecture in FIGURE 2. Surfaces forms of verb are structured in tree hierarchy, as the root in the tree's root is followed with some levels which represent morphology features ended with leave level, which in turn represents the instances of verb forms.

FIGURE 2. RootIT architecture.

Relational database is adopted for storing outputs of the biggest Arabic morphology system – Sarf. Consequently, we have developed a specific sub-tool acts on typed a root *r* into Sarf interface hence registers Sarf outputs (whole surface forms of *r*) inter of database. Sarf's roots are imported from common thesaurus of *Al-ShAh* الصحاح. *Al-ShAh* has 7442 roots of which 5674 trilateral roots and 1829 quadrilateral roots. Overall surface verb forms resulted are 2,775,742 verbs and 514,734 gerunds. By those roots account, we can claim that it comprises all Arabic roots and there was no root left. However, there are some roots that are rarely used or unused at all. To the side of surface root forms storage, some morphology features are enrolled, such features active, passive, tense, mood, aspect, diacritic form. The DBMS of derby is used since being the easiest one for embedding within related systems. The preprocess of document-mode input of raw-data is POS-tagging which is implemented by a POS-tagger.

IV. EXPERIMENT

To assess the accuracy of the RootIT, a series of experiments have been conducted. The effectiveness of the five systems – Khoja *et al.* [20], ArabicStemmer [37], MADAMIRA [13], Buckwalter [21] and our proposed rooter RootIT - has been evaluated and compared in terms of the accuracy of the F-score measure. The data set used in our experiments is extracted from the most popular and standard Arabic corpus (PATB) [38]. The adopted dataset consists of just all verbs with their surface forms scattered in PATB, which are extracted from tagged and stored parsed trees in the enormous system, Stanford CoreNLP. Total number of dataset verbs tagged by embedded tool belong to Stanford called MaxentTagger amount to 7591 surface verbs excluding unsound tagged verbs. Those verbs belong to distinct (1938) roots, (76) quadrilateral root and the rest are trilateral roots.

The evaluation method, that we have adopted, is the commonly used metric and the standard evaluation measure in the IE community. The standard evaluation measures Precision, Recall, and F-measure are used to evaluate the performance of our system to compare the results of the comparative other four systems [39]. The comparative systems cannot find the root for some verbs (a few listed in TABLE 1). The failure in those systems might be due to two factors. The first one is that such systems get satisfied by stem of verbs without diving to root of verb, and the second one is limited rules pattern used by systems against verbs morphology requirements. Although Khoja system, that has obtained efficient accuracy in detecting verbs' roots as well as nouns' roots, it has been unable to recognize some verb patterns that have evasiveness trait with formatted morphology rules in system algorithm of such verbs as تتدلع, تخطئ, اضطلعت, نستمع, …etc.

TABLE 1. Example of verbs rooting results using four compared systems and our system.

Verb	MADAMIRA	Khoja	Arabic-Stemmer	Buckwalter	RootIT
اضطلعوا	اضطلع	-	اضطلع	-	ضلع
يتراءى	تراءى	-	تراء	تراءى	رءي
ساءت	ساء	ساءت	ساء	ساء	سوء
يستخدم	ستخدم	سخد	استخدام	ستخدم	خدم
أفرحي	-	فرح	افرح	-	فرح
لاذ	لاذ	لذي	-	لاذ	لذذ
استسلمنا	إستسلم	سلم	استسلم	إستسلم	سلم

However, due to our root-based approach, RootIT can extract all patterns of a verb and trace correct path root that most match the given verb pattern. Instead of using affixes (prefixes and suffixes) in a list along with formatted template that stop confused toward abundances of Arabic verb morphology.

TABLE 2. Rooting accuracies of the five systems.

	MADAMIRA	Khoja	Arabic-Stemmer	Buckwalter	RootIT
Precision	84.90%	93.55%	64.84%	83.23%	96.26%
Recall	85.88%	95.68%	74.11%	88.78%	98.45%
F-score	85.38%	94.60%	69.16%	85.91%	97.34%

To illustrate that our RootIT is more efficient, we present some results of Arabic verbs' root identification systems mentioned in TABLE 2. TABLE 2 shows the RootIT accuracy of 97.34%. In comparison, we observe that our proposed tool can produce better results as FIGURE 3 illustrated. One of the main points of evaluation is the comparison of RootIT

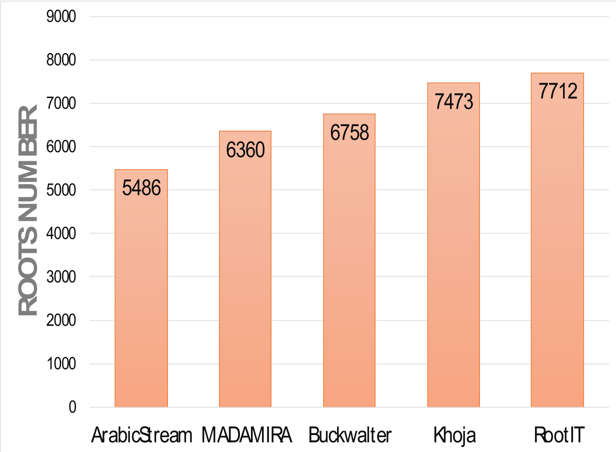

FIGURE 3. Comparison between RootIT and other four systems.

with the most competed systems. Hence, the RootIT results reflects our hypothesis as stated in section I.

As for the weak point of RootIT, it is resulted from some verbs that have a pronoun attached with as suffix like يستقوهم، زوجتك سلمتهم، etc. Nevertheless, it is on the top among competed systems (see FIGURE 3).

V. CONCLUSON

We have developed a detector tool RootIT that enables to identify verb's root exists in Arabic sentence as a first step towards the automated detection of all arguments of verb. We have developed this tool in order to supply linguistic analyzers concerted on verb with its derivations and inflections, in addition to improve the performance of the Arabic morphology systems at rooting verbs module. The method used adopts mapping of the real verb in text against surface verb forms stored in database filled from outcomes of Sarf morphology system. Through the test, PATB dataset verbs are extracted for constructing our evaluation dataset. Four of state-of-the-art rooting system have been adopted for comparing their results with our system results. We have obtained (97.34%) in f-measure as a top accuracy among competed systems. Accompaniment of rest morphology features, which have been already initialed in this model along with root, remains as a future work as well as all derivated nouns owned by root into the database on the goal of identify root of nouns derivations.

REFERENCES

[1] M. S. Basati, "The nominal sentence in Arabic grammar," Tech. Rep., 2018.

[2] W. Wright and C. P. Caspari, *A Grammar of the Arabic Language*. New York, NY, USA: Cosimo, 2011.

[3] N. Kambhatla and I. Zitouni, "Systems and methods for automatic semantic role labeling of high morphological text for natural language processing applications," Google Patent 8 527 262 B2, Sep. 3, 2013.

[4] H. Ishkewy, H. Harb, and H. Farahat. (2014). "Azhary: An arabic lexical ontology." [Online]. Available: https://arxiv.org/abs/1411.1999

[5] N. Habash, "Arabic morphological representations for machine translation," *Arabic Computational Morphology*. Springer, 2007, pp. 263–285.

[6] F. Aqlan, X. Fan, A. Alqwbani, and A. Al-Mansoub, "Improved Arabic–Chinese machine translation with linguistic input features," *Future Internet*, vol. 11, no. 1, p. 22, 2019.

[7] *Sarf*. [Online]. Available: www.alecso.org.tn

[8] M. C. Nouri, *The Three Steps of Action in Arabic—A Phonetic Study Using Computer*. 2018.

[9] A. Abdelali, K. Darwish, N. Durrani, and H. Mubarak, "Farasa: A fast and furious segmenter for arabic," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Demonstrations*, 2016, pp. 11–16.

[10] M. Attia, "Developing a robust Arabic morphological transducer using finite state technology," in *Proc. 8th Annu. CLUK Res. Colloq.*, 2005, pp. 9–18.

[11] A. Amine, L. Bellatreche, Z. Elberrichi, E. J. Neuhold, and R. Wrembel, "Computer science and its applications," in *Proc. 5th IFIP TC Int. Conf. (CIIA)*. Springer: Saida, Algeria, May 2015.

[12] W. Black *et al.*, "Introducing the Arabic wordnet project," in *Proc. 3rd Int. WordNet Conf.*, 2006, pp. 295–300.

[13] A. Pasha *et al.*, "MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic," in *Proc. LREC*, 2014, pp. 1094–1101.

[14] I. A. Al-Sughaiyer and I. A. Al-Kharashi, "Arabic morphological analysis techniques: A comprehensive survey," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 55, no. 3, pp. 189–213, 2004.

[15] E. Benmamoun, "Arabic morphology: The central role of the imperfective," *Lingua*, vol. 108, nos. 2–3, pp. 175–201, 1999.

[16] B. Haddad, A. Awwad, M. Hattab, and A. Hattab, "Associative root–pattern data and distribution in arabic morphology," *Data*, vol. 3, no. 2, p. 10, 2018.

[17] S. Moscati, *An Introduction to the Comparative Grammar of the Semitic Languages: Phonology and Morphology*. Markkleeberg, Germany: Otto Harrassowitz, 1964.

[18] W. Salloum and N. Habash, "ADAM: Analyzer for dialectal arabic morphology," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 26, no. 4, pp. 372–378, 2014.

[19] M. Sawalha and E. Atwell, *Visualisation of Arabic Morphology*. 2012.

[20] S. Khoja and R. J. L. Garside, "Stemming Arabic text," Dept. Comput., Lancaster Univ., Lancashire, U.K., 1999.

[21] T. Buckwalter, "Buckwalter Arabic morphological analyzer version 1.0," Tech. Rep., 2002.

[22] M. N. Al-Kabi, S. A. Kazakzeh, B. M. A. Ata, S. A. Al-Rababah, and I. M. Alsmadi, "A novel root based Arabic stemmer," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 27, no. 2, pp. 94–103, 2015.

[23] R. Alshalabi, "Pattern-based stemmer for finding Arabic roots," *Inf. Technol. J.*, vol. 4, no. 1, pp. 38–43, 2005.

[24] J. Xu, A. Fraser, and R. Weischedel, "Empirical studies in strategies for Arabic retrieval," in *Proc. 25th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2002, pp. 269–274.

[25] M. Ababneh, R. Al-Shalabi, G. Kanaan, and A. Al-Nobani, "Building an effective rule-based light stemmer for Arabic language to improve search effectiveness," *Int. Arab J. Inf. Technol.*, vol. 9, no. 4, pp. 368–372, 2012.

[26] L. S. Larkey, L. Ballesteros, and M. E. Connell, "Light stemming for Arabic information retrieval," *Arabic Comput. Morphol.*. Springer, 2007, pp. 221–243.

[27] K. Taghva, R. Elkhoury, and J. Coombs, "Arabic stemming without a root dictionary," in *Proc. Int. Conf. Inf. Technol., Coding Comput.*, Apr. 2005, pp. 152–157.

[28] R. Sonbol, N. Ghneim, and M. S. Desouki, "Arabic morphological analysis: A new approach," in *Proc. 3rd Int. Conf. Inf. Commun. Technol., Theory Appl.*, Apr. 2008, pp. 1–6.

[29] M. Boudchiche and A. Mazroui, "Improving the Arabic root extraction by using the quadratic splines," in *Proc. Int. Conf. Intell. Syst. Comput. Vis.*, Apr. 208, pp. 1–5.

[30] M. Momani and J. Faraj, "A novel algorithm to extract tri-literal Arabic roots," in *Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl.*, May 2007, pp. 309–315.

[31] A. Al-Omari and B. Abuata, "Arabic light stemmer (ARS)," *J. Eng. Sci. Technol.*, vol. 9, no. 6, pp. 702–717, 2014.

[32] R. Mohammed, "New Arabic stemming based on Arabic patterns," *Iraqi J. Sci.*, vol. 57, no. 3C, pp. 2324–2330, 2016.

[33] K. Darwish, "Building a shallow Arabic morphological analyser in one day," in *Proc. ACL Workshop Comput. Approaches semitic Lang.*, 2002, pp. 1–8.

[34] A. A. Neme, "A fully inflected Arabic verb resource constructed constructed from a lexicon of lemmas by using finite-state transducers," *Revue RIST*, vol. 20, no. 2, p. 13, 2013.
[35] M. Yaseen *et al.*, "Building annotated written and spoken Arabic LRs in NEMLAR project," in *Proc. LREC*, 2006, pp. 533–538.
[36] H. Khafajeh, N. Yousef, and M. Abdeldeen, "Arabic root extraction using a hybrid technique," *Int. J. Adv. Comput. Res.*, vol. 8, no. 35, pp. 90–96, 2018.
[37] *ArabicStemmer*. [Online]. Available: https://www.arabicstemmer.com/
[38] M. Maamouri, A. Bies, T. Buckwalter, and W. Mekki, "The penn Arabic treebank: Building a large-scale annotated arabic corpus," in *Proc. NEMLAR Conf. Arabic Lang. Resour. Tools*, 2004, pp. 466–467.
[39] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," Tech. Rep., 2011.

Authors' photographs and biographies not available at the time of publication.