

Inspiring Technologies and Innovations

June 2025, Volume: 4 Issue: 1

Research Article **Investigation of the Morphological and Root Structure of the Arabic Language Structure Using Python****Abdulmonem AHMED^a, Aybaba HANÇERLİOĞULLARI^b, Ali Rıza TOSUN^c**^aDepartment of, Graduate School of Natural and Applied Kastamonu University, Kastamonu, TÜRKİYE^bDepartment of Physics, Faculty of Science, Kastamonu University, Kastamonu, TÜRKİYE^cDepartment of Philosophy, Faculty of Science and Letters, Kastamonu University, Kastamonu, TÜRKİYEORCID^a: 0000-0001-9816-9717ORCID^b: 0000-0002-9830-4226

Corresponding Author e-mail: aybaba@kastamonu.edu.tr

<https://doi.org/10.5281/zenodo.15735235>

Received : 2.10.2023 Accepted : 31.05.2025 Pages : 1-6

ABSTRACT: In the last years, scientists in computer and software have made successful studies in almost every field. Especially in the classification and comparison of optimization algorithms, metaheuristic optimization techniques can be generally classified as population-based and trajectory-based methods. These algorithms are quite suitable for high-dimensional problems and exhibit strong discovery capabilities. The algorithms use a single solution and iteratively improve it based on mathematical models or heuristics. Arabic is a language of derivation with a very rich derivational morphology, with virtually all words originating from roots through patterns. It is a language with numerous inflections and a complicated morphological structure. Arabic plurals used in modern Arabic are listed. Arabic is becoming a significant topic because of applications for information retrieval and natural language processing. There are two types of Arabic plurals: regular and irregular. The standard form also includes masculine and feminine versions. Stemming is a method used in information retrieval to differentiate between single and plural nouns by eliminating the corresponding affixes from words. In this study, we describe a technique that does not rely on stemming to recognize Arabic plurals. Based on the threefold relationship between word, root, and pattern, we utilize both the word and the pattern to find the correct root and the pattern used to define the type of plural. The algorithm has only been provided the fundamental patterns of plurals, whether regular or irregular; the remaining patterns are constructed programmatically to add the appropriate plugins for each type of plural. Finally, since irregular inclusion and humiliation are not governed by direct laws, we get positive outcomes in both cases. In the situation of irregular inclusion and humiliation, we are largely satisfied. Although generally faster, these methods may experience problems such as premature convergence or being stuck in local optima. Recent research has focused on improving existing algorithms by hybridization, parameter tuning, and incorporating additional mechanisms such as chaos theory and adversarial learning.

KEYWORDS: Python arabic plurals, morphology, root, optimization.

1. INTRODUCTION

The most common problem with metaheuristic algorithms is user-defined parameters. Mathematical models using random numbers negatively affect the global optimum convergence. The LM trap must be managed between exploration and exploitation tendencies to prevent premature convergence. There are many options for working with GM-capable and successful SI and EA algorithms. Birds engage in social interactions to obtain sufficient food. Each individual in the flock represents a solution. To update their position in the search space, for example in PSO, the current speed is multiplied by the current position. The best solution so far has been obtained by the PSO algorithm and it improves with each iteration. The global best position can only be reached by the group members. Subfield of artificial intelligence (AI) called Natural Language Processing (NLP) is focused on understanding and modifying human languages. Each individual human natural language has a unique collection of rules and syntaxes that set it apart from other languages. Writing, translating, and even speaking natural language is challenging since these languages contain inflectional, which adds to the confusion [1]. Humans can interpret meaning from context and have feelings, while computer algorithms and programs struggle to do this and make mistakes [2]. In order to build and improve their job reach tools and application activities in this field, researchers are continually working on these projects. Over time, NLP has developed into a growing field. Examples of NLP fields that have existed and are still in existence include machine translation, text description, sentiment analysis, word segmentation, information retrieval, and text categorization [3]. In NLP applications, the Morphological Analyzer is a significant module that recognizes and examines the structural structure of words in a specific language [6]. The process of generating a new word from an existing one is referred to as morphology. It is common for morphological analysis to be challenging, computationally demanding, and inherently parallel [7]. Arabic information retrieval systems' major objective is to find and retrieve Arabic documents from databases that are crucial to a particular query. In these systems, retrieving documents is done by comparing query and index terms for similarity. Arabic numbers come in solitary, dual, and plural forms. Consequently, the Arabic plural starts at three. Arabic plurals can be either regular or broken (irregular) [8]. Light stemmers swiftly change irregular plurals (broken plurals), which contain infixes in addition to prefixes and suffixes, into their single forms. Regular plurals can be separated from irregular plurals by the addition of a suffix, which is a collection of letters at the end of a word. While broken plurals are problematic, it can be challenging to identify them because of their intricate

patterns [9]. We suggest a technique in this paper that can distinguish between regular and irregular plurals without eliminating any suffixes. Instead, the approach makes advantage of the pattern to identify the word's root before determining its morphology, which in this case is plural.

1.1. Arabic Language Overview

Arabic contains Twenty-eight letters written from right to left make up the word, which can be increased to ninety by including more shapes [10]. Only a small portion of native Arabic words are derived from four, five, or six letter roots, the majority of which are formed from three consonant letters (trilateral roots) [11,12]. It has been challenging to locate standard Arabic text mining algorithms and techniques. Verbs, nouns, and particles are the three divisions of the Arabic language. Arabic has short vowels that are not letters, often known as diacritics. These vowels make grammar checking and root extraction simpler while also reducing the ambiguity of a word's meaning [13]. These patterns are made by attaching affixes to the roots and can be seen as models that adhere to Arabic grammatical standards. To extract various grammatical applications, such as possessives, plurals, definite forms, gender, and so forth, additional affixes might be introduced [14]. Arabic is a language of several roots and both nouns and verbs are descended from a series of roots. The process of "root extraction," also known as "stemming," is a means to locate or recognize the root or stem of any Arabic word [15].

1.2. Arabic Morphology

Arabic's morphological representation is quite intricate due to morphological events like agglutination. Depending on where they are in the word (beginning, middle, end, and separate), letters take on distinct shapes [17]. Morphological analysis is a crucial stage in language processing because of Arabic's intricate morphological structure. It appears that morphology has persuaded traditional Arabic grammarians to classify words only into verbs, nouns, prepositions, and particles. Almost all words in Arabic are formed from roots applying patterns, making it essentially a derivational language with an extremely rich derivational morphology. The bulk of Arabic words have three-letter roots as their foundation. [16, 18]. Arabic is a morphologically complicated language. It is possible to define short vowels, consonantal doubling, and the morpheme using optional diacritics. The absence of these diacritics and the rich morphology of the language lead to a considerable degree of certainty, several Arabic letters are frequently wrongly spelled. Because of these complexities, the Arabic language poses a substantial barrier for NLP [4,19, 20, 21].

1.3. Plurality in Arabic

Arabic numbers come in three different varieties: singular, dual, and plural. As a result, the Arabic plural starts at three. Arabic distinguishes between regular and irregular (broken) plurals [8]. The most exciting case is the one involving the broken plural because it requires a lot of time and effort to solve. There are no set rules for how to produce a broken plural or how to recognize one (such as adding or removing letters from the single form). There are several ways to carry out these processes. Each strategy has advantages and disadvantages [5, 22]. The elimination of affixes, the application of grammatical rules, or even the usage of dictionaries is a few of these strategies.

1.4. Arabic Regular Plural

The Arabic regular plural is formed by adding no additional letters to the stem. It's also important to remember that the pattern of an Arabic word consists of three letters, which are typically represented by the letters "ع", "ف", and "ل". Arabic nouns and adjectives can be classified as either masculine or feminine. Because of this, plurals are formed using various inflectional suffixes with little to no internal shift [8]. There are two types of this plural component, as follows:

1.5. Arabic Masculine Regular Plural

The usable nouns in male standard plurals are the sensible masculine proper nouns, whose bases do not terminate in a vowel letter, and the sensible masculine adjectives. The masculine regular form gains the suffixes "ون" or "ين" [8,9].

1.6. Arabic Feminine Regular Plural

This common Arabic plural can be used to refer to a broad variety of noun categories, including both human and nonhuman individuals, as well as adjectives. The typical feminine plurals are formed with the suffix "ات"; it should be noted that when the stem ends in the letter "ي", the suffix takes its place [8,9].

1.7. Arabic Irregular Plural

In Arabic, there are a number of specific requirements that must be followed in order to construct irregular plurals, which lack a clear structure [8]. It's important to note that the Arabic language has a high degree of inflectional structure, with more than 85% of all words having trilateral origins. Arabic verbs and nouns are typically derived from a group of roots [23]. Broken plurals make up 10% of documents in large Arabic corpora, while plurals make up 41% [9]. Broken plural identification is a serious issue for light-stemming algorithms created for applications like information retrieval. Arabic has tight rules for how to produce irregular plurals, which lack a clear structure [8]. A root is the most fundamental word in phonetics, used as a base for the addition of suffixes or affixes to produce various derivatives like verbs, adjectives, and nouns. It's important to note that approximately 85% of Arabic words have trilateral origins, making it a highly inflectional language [7].

2. MATERIAL AND METHOD

A few studies have concentrated solely on the Arabic plural, despite the fact that a great deal has been done in the area of Arabic morphological analysis and generation using a range of techniques and at various levels of linguistics. These few studies focus on the removal of singular nouns from their plural forms as well as the derivation of the plural from the singular or the base. In this section, we discuss a few of these studies in order to better grasp their methods for morphological study of Arabic as a whole. In the Table 1, we will discuss some of the prior studies on the subject of identifying verbs in various ways and at various periods in the rest of this section of the study.

Table 1. Plural Recognizing Related Works

Years	Author(s)	Discussion
2013	Alexis Amid Neme and Eric Laporte	They proposed a model that is implemented and takes broken plurals into account. The model uses a lexicon of terms to do direct morphological analysis of Arabic text rather than using morphophonological criteria. by defining vowel quantity and neglecting vowel quality to streamline the taxonomy of singular patterns. Without deep roots or morphophonological or orthographical principles, root alternations and orthographical variants are recorded independently from patterns and in a factual manner [24].
2014	Abduelbaset Goweder, Massimo Poesio, Anne De Roeck and Jeff Reynolds	They show that irregular plural recognition in modern standard Arabic is a difficult problem for information retrieval and language engineering applications. They developed a number of methods for detecting irregular plurals and tested them. The irregular plural detection component was integrated into a new light-stemming algorithm that conflates both regular and irregular plural with their singular forms [16].
2016	Ali Shafah, Abduelbaset Goweder, Samira Eshafah and Ahmed Rgibi	They proposed a method for identifying broken plurals in Arabic without the need for stemming. they used the decision tree framework (WEKA J48) to construct a classifier (model). An unknown test set is used to evaluate the constructed classifier. The findings show that a promising broken plural recognizer for natural language processing applications could be developed and implemented [9].

The field of Arabic language morphology has seen a wide range of studies and analysis. Many of these studies employ either applying either grammatical rules or eliminating affixes, or even using both of them together. The research methods and algorithms used to create a morphological analyzer are different. The majority of research in the field of plural has concentrated on irregular plural. We will present an approach in this paper that uses the pattern to find the root and then determine the form of plural from the pattern used. In This paper, we introduce a technique that can extracts, roots directly without any affix removal, it just matches the words to extract its root by a pattern with the same number of letters, and we will discuss in more detail the ideas of our algorithms in section four.

3. RESULTS AND DISCUSSION

Regular and irregular Arabic plurals are the two categories. The common form contains both masculine and feminine forms. A crucial and challenging issue that needs to be solved is the detection of plurals. In this work, we have presented a method for identifying Arabic plurals without removing affixes or using grammar rules, but rather by relying on the three-way relationship between word, root, and pattern to determine the word's structure. Linguists consult the Holy Qur'an and take the Qur'an's requirements for proper grammar into consideration since they believe that the Qur'an is the language of the book. With 6144 words, Surat Al-Baqarah is the largest surah in the Qur'an, thus we used it to test our technique. Although the letters of the Noble Qur'an include diacritical marks, our algorithm works with letters without them since the Arabic used today, whether in printed or electronic books or even online pages, lacks diacritical marks. An algorithm for detecting Arabic plural was presented. In this method, the pattern is used to extract the root, and the second step is to determine how to determine the plural from the pattern that was used to extract the root. After the text has been preprocessed, which removes all extraneous words from the original text such as stop words, vowels, digits, and non-text characters, these two processes are carried out. The fundamental concept is to consider the connections between the three components, word, pattern, and root. In our suggested approach, we have utilized these relations and translated them. The steps of this method are shown in Figure 1.

```

1. Algorithm RecognizeArabicPlular(ArabicText)
2.   // input text as Arabic Words.
3.   // output: List of Arabic Plular.
4.   for Patt in PatternsList:
5.     Root = ''
6.     for each char in Patt:
7.       if char in ['F','A','L']:
8.         root = root + char
9.       end if
10.    end for
11.    word = ''
12.    for each char in Patt:
13.      if char in ['F','A','L']:
14.        word = word + root(char)
15.      end if
16.    end for
17.    if word == WORD:
18.      print('word is Plular frm type Patt')
19.      exit for
20.    end if
21.  end for
23. end Algorithm

```

Figure 1. Algorithm to identify Arabic plural

The basic patterns for plurals as well as affixes that can be added to these patterns have been provided to the algorithm. The algorithm itself will produce all additional patterns that are employed in opposition to all words. The patterns used to identify masculine regular plural are shown in Tables (2), (3), and (4), respectively, whereas the patterns for irregular plural are shown in Table (3).

Table 2. Patterns for masculine regular plural

اتفعل	فاعل	مفعل
اتستفعل	فعال	مستفعل
افتعمل	فعل	يتفعل
افعل	فعلان	يستفعل
تتفعل	فعلي	يفتعل
تفاعل	فعول	يفعل
تفعال	فعليل	
تفعل	متفعل	

Table 3. Patterns for feminine regular plural

مفتعل	فعال	اقتعال
مفاعل	فعل	افعال
مفتعل	فعلاو	تفاعل
مفعّل	فعلي	تفعيل
مفعول	فعيل	فاعل

Table 4. Patterns for irregular plural

فعلهم	فعلياكم	افو عليكم	افعلاء	فعال
فعلوا	فافعلنا	بفعلهم	فعلهم	افعال
فعلناهم	مفعلهم	افعلناكم	افعالهم	فعلة
الفواعل	يفعلونهم	افعلوا	افو علها	افعلنا
بفعلهم	يفعلهم	فعلتكم	فعولها	فعول
الافعال	تفعلتم	يفعلونكم	مفاعيل	تفاعيل
الفعالية	افعلواهم	فعلتم	فعولهم	مفاعل
فافعلوا	فعلة	بافعالكم	افعلنا	فعلاء
فعلتنا	فعولكم	فافعلوا	افعالكم	فعلاءكم
بفعلايهم	ليفعلوا	فعلونا	الافعل	فعالكم

4. CONCLUSION

Morphology is a subfield of linguistics that studies the internal structure of words. It examines how affixation affects word creation as well as word origins and pattern characteristics. Case, gender, number, tense, person, mood, and speech are just a few of the properties that can be impacted by inflection [16]. The remainder 2133 words after excluding diacritics, numerals, stop words, and repeated words. There was a 98 percent recognition rate for the masculine regular plural, a 97 percent identification rate for the feminine regular plural, and a 90 percent identification rate for the broken plural. We did not use a dictionary or lexicon, and our whole process did not make use of any grammatical rules. One advantage of what we've done is that it processes data more slowly than other algorithms, which are frequently composed of intricate IF statements. We think that by mixing many algorithms simultaneously, we can get superior outcomes. For example, if we apply grammatical rules and use language dictionaries in our algorithm, we can achieve more accurate results.

REFERENCES

- [1] Q. Yaseen and I. Hmeidi, "Extracting the roots of Arabic words without removing affixes," *Journal of Information Science*, vol. 40, no. 3, pp. 376–385, 2014.
- [2] R. Kanaan and G. Kanaan, "An improved algorithm for the extraction of trilateral Arabic roots," *European Scientific Journal*, vol. 10, no. 3, 2014.
- [3] M. El-Defrawy, N. A. Belal, and Y. El-Sonbaty, "An efficient rank based Arabic root extractor," *2017 Intelligent Systems Conference (IntelliSys)*, pp. 870–878, 2017.
- [4] A. Abu-Errub, A. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Arabic roots extraction using morphological analysis," *International Journal of Computer Science Issues (IJCSI)*, vol. 11, no. 2, p. 128, 2014.
- [5] S. Ellouze, K. Haddar, and A. Abdelwahed, "NooJ disambiguation local grammars for Arabic broken plurals," *Proceedings of the NooJ 2010 International Conference*, pp. 62–72, 2010.
- [6] I. Damaj, M. Imdoukh, and R. Zantout, "Parallel hardware for faster morphological analysis," *Journal of King Saud University - Computer and Information Sciences*, vol. 30, no. 4, pp. 531–546, 2018.
- [7] W. Etaiwi and A. Awajan, "Graph-based Arabic NLP techniques: A survey," *Procedia Computer Science*, vol. 142, pp. 328–333, 2018.

- [8] A. Himmah and R. Wahyudi, "A contrastive analysis of Arabic and English noun plural markers," *PAROLE: Journal of Linguistics and Education*, vol. 4, no. 2, pp. 72–87, 2014.
- [9] A. Shafah, T. Ould-Brahim, A. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Irregular Arabic plurals recognition without stemming," In: 2016 4th International Conference on Control Engineering & Information Technology (CEIT). IEEE, pp. 1–6, 2016.
- [10] T. Kanan, Y. Kanaan, A. Alsmadi, and M. Hawashin, "Arabic light stemming: A comparative study between p-stemmer, khoja stemmer, and light10 stemmer," 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 511–515, 2019.
- [11] I. A. Asadi and A. Khateb, "Predicting reading in vowelized and unvowelized Arabic script: An investigation of reading in first and second grades," *Reading Psychology*, vol. 38, no. 5, pp. 486–505, 2017.
- [12] M. Goudjil, M. Ennaji, and A. Guessoum, "Arabic text categorization using SVM active learning technique: An overview," 2013 World Congress on Computer and Information Technology (WCCIT), pp. 1–2, 2013.
- [13] A. H. Krea, A. S. Ahmad, and K. Kabalan, "Arabic words stemming approach using Arabic WordNet," *International Journal of Data Mining & Knowledge Management Process*, vol. 4, no. 6, p. 1, 2014.
- [14] M. N. Al-Kabi, A. A. Al-Ayyoub, Y. Jararweh, and A. A. Huneiti, "A novel root based Arabic stemmer," *Journal of King Saud University - Computer and Information Sciences*, vol. 27, no. 2, pp. 94–103, 2015.
- [15] A. Gowedar, K. R. Beesley, and L. Karttunen, "Identifying broken plurals in unvowelised Arabic text," *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 246–253, 2004.
- [16] M. Gridach and N. Chenfour, "Developing a new approach for Arabic morphological analysis and generation," *arXiv preprint arXiv:1101.5494*, 2011.
- [17] R. Sonbol, N. Ghneim, and M. S. Desouki, "Arabic morphological analysis: A new approach," 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, pp. 1–6, 2008.
- [18] W. Salloum and N. Habash, "ADAM: Analyzer for dialectal Arabic morphology," *Journal of King Saud University - Computer and Information Sciences*, vol. 26, no. 4, pp. 372–378, 2014.
- [19] A. M. Bashir, M. Q. Abughofa, A. M. Kanaan, and M. Z. Rehman, "Implementation of a neural natural language understanding component for Arabic dialogue systems," *Procedia Computer Science*, vol. 142, pp. 222–229, 2018.
- [20] T. Kanan, A. Alsmadi, and M. Hawashin, "A review of natural language processing and machine learning tools used to analyze Arabic social media," 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), pp. 622–628, 2019.
- [21] E. Issa, "An OpenNMT model to Arabic broken plurals," *Proceedings of the First International Workshop on Language Cognition and Computational Models*, pp. 22–30, 2018.
- [22] A. Alnaied, M. Elbendak, and A. Bulbul, "An intelligent use of stemmer and morphology analysis for Arabic information retrieval," *Egyptian Informatics Journal*, 2020.
- [23] A. A. Neme and E. Laporte, "Pattern-and-root inflectional morphology: the Arabic broken plural," *Language Sciences*, vol. 40, pp. 221–250, 2013.