Using Roots and Patterns to Detect Arabic Verbs without Affixes Removal

Abdulmonem Ahmed[1], Aybaba Hançrlioğulları[2] , Ali Rıza Tosun[3]

1Department of Material Science and Engineering, Graduate School of Natural and Applied Kastamonu University, 37150 Kastamonu /Turkey monemalish@gmail.com ,ORCID:0000-0001-9816-9717
2Department of Physics, Faculty of Science and Letters, Kastamonu University, 37150 Kastamonu /Turkey
3Department of Philosophy, Faculty of Science and Letters, Kastamonu University, 37150 Kastamonu /Turkey

ABSTRACT

Morphological analysis is a branch of natural language processing, is now a rapidly growing field. The fundamental tenet of morphological analysis is that it can establish the roots or stems of words and enable comparison to the original term. Arabic is a highly inflected and derivational language and it has a strong structure. Each root or stem can have a large number of affixes attached to it due to the non-concatenative nature of Arabic morphology, increasing the number of possible inflected words that can be created. Accurate verb recognition and extraction are necessary nearly all issues in well-known study topics include Web Search, Information Retrieval, Machine Translation, Question Answering and so forth. in this work we have designed and implemented an algorithm to detect and recognize Arbic Verbs from Arabic text.The suggested technique was created with "Python" and the "pyqt5" visual package, allowing for quick modification and easy addition of new patterns. We employed 17 alternative patterns to represent all verbs in terms of singular, plural, masculine, and feminine pronouns as well as past, present, and imperative verb tenses. All of the verbs that matched these patterns were used when a verb has a root, and the outcomes were reliable. The approach is able to recognize all verbs with the same structure without requiring any alterations to the code or design. The verbs that are not recognized by our method have no antecedents in the Arabic roots. According to our work, the strategy can rapidly and precisely identify verbs with roots, but it cannot be used to identify verbs that are not in the Arabic language. We advise employing a hybrid approach that combines many principles as a result.

Keywords:
Verb; Arabic; root; pattern

1. INTRODUCTION

One of the top languages spoken worldwide, Arabic is one of the recognized languages of the United Nations and has millions of original speakers [1], and one of the most popular languages on the Internet [2]. Arabic is a linguistic variation well renowned for its deep and rich morphology [5]. The computerized translation from and into Arabic has become much more difficult because of the addition of Arabic morphology to other language components [1, 3, 4]. Arabic is renowned for its inherent ambiguity, with a single hword indicating a variety of meanings [6, 7], furthermore, vocalizations can be used when writing Arabic (diacritics). The vowel letter does not have any diacritical marks, and

each word letter in Arabic has a sign call (haraka), which literally means "motions." [8, 9] The possible candidates are Fathah, Kasrah, Dammah, Sukun, Shaddah, and Tanween. Arabic has a number of morphological characteristics that make error diagnosis and feedback difficult [10, 11, 12]. The need to evaluate Arabic morphology and its application in automated analysis systems are two issues that the Arabic language is currently facing. As a result, word analysis and information retrieval are crucial steps in many NLP applications [13, 14]. Dealing with the internal structure of words is usually referred to as morphology; this aspect analysis helps analyze the word constitution, including affixation, roots, and pattern [15, 16]. Inflectional or derivational are the two fundamental ideas used to categorize morphology. A stem with a known formation is given inflectional treatment. The grammatical class of the word, such as noun, verb, and so on, remains unaffected but, Inflection can affect gender, number, tense, person, and voice.

On the other hand, derivational morphology adds a group of morphemes to a given word that may change its syntactic category [13, 17]. It is difficult to distinguish between these two classifications, and it varies depending on the language [13]. The Rule-based and statistically based approaches to Arabic NLP are the two most used methods. The rule-based methodology is based on manually generated morphological or grammatical rules that were created by linguists. Arabic described as having an interwoven root and pattern. The consonantal root of an Arabic verb is made up of three or four consonants that are often constant across all of its conjugated forms; A conjugated form's remaining details are referred to as its pattern. The patterns change the root's meaning to produce a number of related words [20]. Without compromising the sense, a fragmented text should be paraphrased, some systems aim to extract a verb conjugation, while others only need the verb's surface form [21]. The study of Arabic morphology has been the subject of several researches. They employ three strategies in these works, the first approach using symbols, the second approach using statistics, and third is hybrid approach. The Symbolic method relies on stripping away all prefixes and suffixes in an effort to get at the Arabic surface word's source,   In statistical methodology, the probability that a prefix, suffix, or template will exist in a corpus of surface words is calculated, While the hybrid approach combines morpheme

combination rules with statistics [25]. The majority of Arabic texts that are still in existence are created in the official Modern Standard Arabic Text language (MSA), but as more people use social media sites like Facebook, Google Moderator, and Twitter, the quantity of informal, colloquial Arabic texts grows. The majority of Arabic texts that are still in existence are written in the formal Modern Standard Arabic Text (MSA), but as more people use social media sites, the quantity of informal, colloquial Arabic texts grows [26]. Arabic ordinary speech is composed of particles, articles, conjunctions, pronouns, adjectives, adverbs, and adjective parts. The verb is a word that expresses a state of affairs or the occurrence of an action at a specific moment; it illustrates action that is tied to time and can take place in the present, past, or future [30]. We have devised and implemented a morphological analyzer that is independent of the stems dictionary and does not execute the morphological analysis using the rules in order to deal with these problems. The most of studies on Arabic language processing employ methods for removing affixes from word, the approach we use in this work does not require grammar or machine learning applications. Instead, our approach repeatedly compares the word with patterns of equal length until it is found. It begins with the word's root and uses the pattern to determine whether or not the word is a verb.

2. BACKGROUND

Arabic is divided into three primary varieties: Classical Arabic, used in literary materials and the Quran; Modern Standard Arabic, used in formal speeches and writing; and Arabic Dialects, used in casual conversations and day-to-day interactions [2]. Due to the morphological characteristics of Arabic sentences and words, the syntax and semantics of Arabic texts are difficult [31]. Both the grammar and the semantics of the sentence are complicated. It also includes an enormous number of vocabularies, such as word origins used as nouns or verbs, synonyms, and antonyms [32]. Due to its complexity in form and meaning, Arabic sentence construction depends on the verb morphology. Arabic verbs come in two voices (active and passive), and three tenses (perfect, imperfect, and imperative). Using well-behaved morphological principles, verbs can be conjugated in a variety of tenses, voices, and moods. The abnormalities are brought on by some root consonants' phonological restrictions [11]. Arabic is primarily a derivational language, meaning that the majority of its words are constructed using a variety of morphological "patterns" that combine several base words (roots). Around of (9464) Arabic roots can be found, most of these roots are trilateral, and the remaining are quadrantile roots. only (2214) of the aforementioned roots are being used often. The experimental study conducted that only (130) of these are regularly utilized. Words in Arabic can be verbs, nouns, or particles. Arabic verbs fall into three categories: imperative, present, and past. A limited number of morphological forms

created by preceding researchers and experts effectively encode the morphological characteristics for each category. The behavior of verbs when coupled with pronouns, which frequently occurs in Arabic prose, is also reflected in these forms. Thus, storing all pertinent morphological forms and their properties is one of the most effective methods for encoding verb knowledge [27].

RELATED WORKS

Natural language processing research is a major topic in software development and research. Arabic contains a complicated set of morphological rules, a rich feature system, and a combination of template and affixation morphemes [23]. Because Arabic verb morphology has such a vast variety of form and meaning, it is essential to the creation of an Arabic sentence. Arabic verbs can be derived from triangular or quadrangular roots using one of the most common techniques [11]. In Arabic, the verb is frequently used to convey a past general or particular circumstance, an assertion, or a recollection of a past state or occurrence indicated by the primary Verb [30]. 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. Verb Detection related works

Years	Author(s)	Discussion
2020	Mohamed Tahar Ben Othman, Mohammed Abdullah Al-Hagery, And Yahya Muhammad EL Hashemi	A model for understanding an Arabic sentence is presented to identify Arabic verbs, Furthermore, a novel morphological model based on regular expressions is established. All three-letter Arabic verb roots were organized in a hash table. The syntax of Arabic sentences was derived and examined using a collection of principles that make up Arabic grammar. Their regular expression engine can identify about 87% of the verbs that are displayed [32].
2022	Raghad Alkhudair and Mohammad Aljutaily	They give a detailed examination of the pattern of trilateral perfect passive verbs. They demonstrate that the pattern shows many diachronic transitions among speaker generations, which suggests that one attribute of a language is changeability, according to the data. It is hoped that these results would significantly advance our understanding of the literature in general and the passive construction in particular [19].
2019	Teguh Ikhlas Ramadan, Moch Arif Bijaksana, and	Describe a technique for recognizing verb pattern types in classical Arabic. For those who want to understand Arabic from a morphological standpoint and find it essential to process material derived from the

	Arief Fatchul Huda	Holy Qur'an in particular, it makes studying Arabic easier. They show that comprehension of verbs is the initial step in comprehending Arabic morphology. The purpose of this method is to comprehend how the rules for reading patterns with prefixes, diacritical markings, and verb suffixes operate [22].
2017	Samia Ben Ismail, Sirine Boukédi, and Kais Haddar	For the Arabic language, they developed a transformation system that generates derivational versions of each variety of Arabic verb with head-driven grammar. In order to validate them with the platform for constructing linguistic knowledge, the proposed type hierarchy is used to formulate the imagined derived forms in Type Description Language. Most of the derivational forms are produced by the rules that have been introduced to Linguistic Knowledge Building, which also provides solid results quickly [23].

PREPROCESSING

Pre-processing is a method for enhancing a text by eliminating superfluous elements [28]. It's a crucial step in improving the effectiveness of algorithms. The texts must be free of punctuation, diacritics, numerals, and non-Arabic letters [6, 36]. In the area of Arabic word preprocessing, a variety of procedures are frequently applied.

Tokenization

Is a method of segmenting text so that it separates each word from the next by using the initial space. it is the process of swapping out sensitive data with one-of-a-kind identification symbols that keep all of the data's necessary information while without jeopardizing its security [35].

Normalization

Removal of unwanted letters, punctuation, and non-letters from a text is known as normalization. The letters are being returned to their former positions through this process. The normalization process consists of the following steps: remove non-letters, remove punctuation, and remove diacritics [36, 37].

Stop words Removal

Words with minimal meaning are included in a stop word list. they are merely used to describe syntactic rules [18]. Due to their relatively high frequency, they have the potential to influence the retrieval process. Additionally, they tend to minimize the effect of frequency variations amongst the unusual words, which finally influence the weighting procedure [38]. The document length is altered by eliminating the stop words, which has an impact on the weighing procedure. Due to their nature and the lack of

meaning they convey, they also influence the efficiency of text processing, which could lead to a significant quantity of ineffective processing [29].

1. PROPOSED MODEL

We will go through some key ideas before describing the technique used to locate Arabic verbs so that you can understand it before we get into the specifics of our work. The roots of the Arabic language form its basis. A word's root is its most fundamental form, the part of the word that remains after all affixes have been eliminated, and cannot be further explored without losing its identity [33]. Simple nouns and verbs are produced by combining these roots with a range of vowel patterns, which affixes can be attached to produce additional complex derivations and forms [24]. Patterns play a big role in Arabic morphology and lexicography. They are made through the processes of vocalization and affixation. Multiple patterns are produced from the base pattern, and each root can be coupled in a canonical way with orthographically distinct patterns to create new words [33]. In order to extract the root using the pattern and then apply the pattern to identify the word and ascertain the verb, we use the previously described notions of the root, pattern, and word. Pre-processing which is a set of fundamental procedures in most NLP applications. In this work, we have design these preprocessing functions as following, in Figure (1) the function of normalization and diacritics removal, which is showing how to replace some letters with their original letter, and shows how to remove diacritics that are not important here in our work. The procedure of separating and deleting repeated words is a very significant process in the processing of natural languages, As a result, we have provided in Figure (2) below the mechanism of how this function operates. In Figure (3), the function to remove stop words, which was previously prepared in a list or text file, is explained so that it is not processed as previously explained in the preceding sections. A list of clean Arabic words that can be utilized for verb recognition tests or other NLP exercises is the output of the third function.

```
Function Normalize(text):
    import regex  // Regular Expression
    noise = re.compile("""ّ     | # Tashdid
                          َ    | # Fatha
                         ً    | # Tanwin Fath
                         ُ    | # Damma
                         ٌ    | # Tanwin Damm
                         ِ    | # Kasra
                         ٍ    | # Tanwin Kasr
                         ْ    | # Sukun
                         ـ      # Tatwil/Kashida
                      """, re.VERBOSE)
    text = re.sub(noise, '', text)
    text = re.sub("ٱ", "[آأإ]", text)
    text = re.sub("ي", "ى", text)
    text = re.sub("ء", "ؤ", text)
    text = re.sub("ء", "ئ", text)
    Return text
End Function
```

Figure (1) Normalization function

```
Function Tokenization(text):
    NewList = []              // new list
    wordList = []             // new list
    wordList = text.split()   // tokonization
    For each word in wordList:
        if word not in wordLis:
            NewList.append(word)
    End For
    Return NewList
End Function
```

Figure (2) Tokenization function

```
Function RemoveStopWords(wordLis):
    Stopwordlist = Readfile('stopwords.txt)
    NewList = []              // new list
    For each word in wordLis:
        if word not in Stopwordlist:
            NewList.append(word)
    End For
    Return NewList
End Function
```

Figure (3) Stop words removal function

The preprocessing procedure is finished, and the data is now available for usage. The method in Figure (4) calls the three previous functions sequentially, reads each word in the list, and then compares it to a pattern with the same amount of letters. A temporary root is then created from the word by comparing the basic letters in the pattern to their positions within it. This temporary root is used for the same pattern; if this word is not produced, the temporary root and pattern are ignored, the same word has been compared with the remaining patterns, and it is not stopped until the word and root have been obtained or the end of patterns has been reached; if the temporary root and pattern produce the same entered word, the proper root is chosen, and the pattern is determined. The pattern will be chosen and selected by the algorithm, which will then determine if it is a pattern for verbs or not. As a result of the fact that it is frequently used to refer to verbs, adjectives, and other nouns.

```
Algorithm to determine Arabic Verbs ()
    // input:  Arabic Document or text
    // output:  List of determined Names
    Doc: = ReadArabicDocument ();
    Doc: = Normalize (Doc);
    Doc: = Tokenization (Doc);
    Doc: = RemoveStopWords (Doc);
    PatternList = ReadFile('StopWords.txt')
    VerbList   = ReadFile('verbs.txt')
    verbs = []
    For Each word in Doc:
        For Each Patt in PatternList:
            Root: = Compare (word, Patt);
            Tmp: = Compare (Root, Patt);
            If word == Tmp Then:
                If Patt in VerbList then:
                    verbs.append(word)
                    Exit For
                End IF
        End For
    End For
    Return verbs
End Algorithm;
```

Figure (4) Detecting Arabic verb Algorithm in pseudocode

2. METHODOLOGY

The broad plan and justification for the research project are referred to as the methodology. In order to create a strategy that meets objectives, it entails understanding the approaches currently being employed in the field as well as the ideas or guiding concepts that underpin them. The process of putting a plan or design phase into action is called implementation. Therefore, if something is to genuinely happen, action must occur before any initial idea. For an implementation process to be effective, numerous tasks between various stages must be finished. We work hard to use professional techniques and tried-and-true processes in order to get the greatest results. We constructed preprocessing stages and presented an algorithm using the Python programming language because it is currently one of the most widely used programming languages in a variety of industries and domains, that also uses natural language processing and machine learning. The fact that Python is a free and open-source programming language is another factor. As a result, several of the functions and processes we developed can be easily utilized in other applications, with the potential for future development and expansion. Finding the root is the morphological analyzer's most crucial role. In this field, the majority of algorithms either remove affixes or follow morphological and grammatical norms. The meaning lies at the core of the pattern, which encompasses all morphological combinations. We'll go through some basic algorithms designed to take a word's root and identify its pattern of correspondence. In this paper, we offer a root extraction algorithm that is composed of three straightforward phases, each of which is referred to as a stage of a sub-algorithm. These steps are carried out following the text's preparation, during which all undesirable terms such as stop words, vowels, digits, etc. are eliminated.

3. DISCUSSION

Our root-based approach overcame the ambiguity of Arabic verbs by using the link between word-root-pattern and name discovery and recognition. The primary goal of this work is to investigate the connections between the three elements "word", "pattern" and "root". The Arabic term is concentrated on the connection between the pattern and the root. If we follow the right pattern, separating the word's root from it is not a challenging problem. Determining the right pattern is crucial. If they are a part of the pattern or not, the affixes occasionally cannot be continuous. We can immediately access the roots without removing affixes, which speeds up the process and avoids eliminating the word's starting letters. The presence, arrangement, and comparison of the letters inside the word as well as the letters within the pattern letters besides the time required for the removal process itself, the removal of affixes can occasionally result in blunders by omitting some of the original word's fundamental letters. Affixes are added or subtracted from the pattern rather than being added or removed, and the result, the root, is then reviewed after

comparing the same-length words and patterns. The obtained root is then used to retrieve the initial word. In a reverse approach because the root is necessary and is assumed to be true if the original word is obtained. All these elements make it challenging to carry out the required procedures when dealing with natural languages. In such exceptional circumstances, by determining how the letters are ordered in both the word and the pattern, assigning values or weights, and choosing the most valuable or weight as the proper root, we can easily compare different patterns to determine which ones are most appropriate.

4. CONCLUSION

There is a large degree of word overlap and word migration into Arabic. One root can be many terms when it interacts with a pattern. In this work, we introduced a pattern-based method and morphological features. Our approach primarily focuses on the word's structure rather than the existence or placement of individual letters in the word. The lack of distinction between the basic letters of the word and the extra letters is one of the most difficult problems that researchers in this subject must deal with. These modifications found in the majority of natural languages, and they serve an important purpose in converting speech into a collection of several temporal conjugations, including past, present, future, and others. We had presented a method for identifying Arabic verbs in texts from electronic news sources. This work makes a contribution by offering a method for verb recognition. The suggested method compares two original words and a pattern with the same number of letters instead of removing any affixes at all from any of its phases. An arbitrary candidate root is produced as a result of this comparison and is then utilized to produce the same word as the one that was extracted from. We have put the suggested algorithm to the test using various Arabic text files in order to assess it. In our tests, we discovered that for terms with Arabic origins, verb recognition is extremely accurate.

REFRENCES

[1] Ameur, Mohamed Seghir Hadj, Farid Meziane, and Ahmed Guessoum. "Arabic machine translation: A survey of the latest trends and challenges." Computer Science Review 38 (2020): 100305.

[2] Guellil, Imane, et al. "Arabic natural language processing: An overview." Journal of King Saud University-Computer and Information Sciences 33.5 (2021): 497-507.

[3] Alkhatib, Manar, and Khaled Shaalan. "The key challenges for Arabic machine translation." Intelligent Natural Language Processing: Trends and Applications. Springer, Cham, 2018. 139-156.

[4] Habash, Nizar, and Fatiha Sadat. "Arabic preprocessing schemes for statistical machine translation." Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers. 2006.

[5] Tahat, Abdelrazzaq. "Morphological Patterns of Personal Naming Practice: A Case Study of Non-Concatenative Arabic Language." (2020).

[6] Zayed, O. H., & El-Beltagy, S. R. (2012, May). Person name extraction from modern standard Arabic or colloquial text. In 2012 8th International Conference on Informatics and Systems (INFOS) (pp. NLP-44). IEEE.

[7] Abdallah, S., Shaalan, K., & Shoaib, M. (2012, March). Integrating rule-based system with classification for arabic named entity recognition. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 311-322). Springer, Berlin, Heidelberg.

[8] Mansouri, A., Affendey, L. S., & Mamat, A. (2008). Named entity recognition using a new fuzzy support vector machine. IJCSNS, 8(2), 320.

[9] Chiu, J. P., & Nichols, E. (2016). Named entity recognition with bidirectional LSTM-CNNs. Transactions of the association for computational linguistics, 4, 357-370.

[10] Alasmari, Jawharah, Janet CE Watson, and Eric Atwell. "A comparative analysis of verb tense and aspect in Arabic and English using Google Translate." International Journal on Islamic Applications in Computer Science and Technology 5.3 (2017): 9-13.

[11] Shaalan, Khaled, Marwa Magdy, and Aly Fahmy. "Analysis and feedback of erroneous Arabic verbs." Natural Language Engineering 21.2 (2015): 271-323.

[12] Ray, Santosh K., and Khaled Shaalan. "A review and future perspectives of arabic question answering systems." IEEE Transactions on Knowledge and Data Engineering 28.12 (2016): 3169-3190.

[13] Muhammed, Muhammed H., Bassim M. Salih, and Omer K. Jasim. "An Emerging Standard Miniaturization in Arabic Morphological Analysis." 2018 1st Annual International Conference on Information and Sciences (AiCIS). IEEE, 2018.

[14] Mourad Gridach and Noureddine Chenfour, Developing a New Approach for Arabic Morphological Analysis and Generation, International Joint Conference on Natural Language Processing,Vol.2, 2011.

[15] Mohammed Attia, Pavel Pecina, Antonio Toral, Lamia Tounsi and Josef van Genabith, An Open-Source Finite State Morphological Transducer for Modern Standard Arabic, FSMNLP '11 Proceedings of the 9th International Workshop on Finite State Methods and Natural Language Processing, 2011.

[16] AbdelRahim A. Elmadany, Sherif M. Abdou and Mervat Gheith, A Survey of Arabic Dialogues Understanding for Spontaneous Dialogues and Instant Message, International Journal on Natural Language Computing (IJNLC) Vol. 4, No.2, 2015.

[17] A. FARGHALY and K. SHAALAN, Arabic Natural Language Processing: Challenges and Solutions, ACM Trans. Asian Language Information Processing, Vol. 8 No. 4, 2009.

[18] Alshalabi, H., Tiun, S., Omar, N., & Albared, M. (2013). Experiments on the use of feature selection and machine learning methods in automatic malay text categorization. Procedia Technology, 11, 748-754.

[19] Alkhudair, Raghad, and Mohammad Aljutaily. "A prosodic morphophonological analysis of the trilateral perfect passive verbs in Qassimi Arabic." Heliyon 8.8 (2022): e10008.

[20] Azman, Bakeel. "Root identification tool for Arabic verbs." IEEE Access 7 (2019): 45866-45871.

[21] Kambhatla, Nandakishore, and Imed Zitouni. "Systems and methods for automatic semantic role labeling of high morphological text for natural language processing applications." U.S. Patent No. 8,527,262. 3 Sep. 2013.

[22] Ramadhan, Teguh Ikhlas, Moch Arif Bijaksana, and Arief Fatchul Huda. "Rule based pattern type of verb identification algorithm for the holy qur'an." Procedia Computer Science 157 (2019): 337-344.

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

[24] Boudchiche, M., & Mazroui, A. (2018, April). Improving the Arabic root extraction by using the quadratic splines. In 2018 International Conference on Intelligent Systems and Computer Vision (ISCV) (pp. 1-5). IEEE.

[25] Yousfi, A. "The morphological analysis of Arabic verbs by using the surface patterns." IJCSI International Journal of Computer Science Issues 7.3 (2010): 11.

[26] Hegazi, M. O., Al-Dossari, Y., Al-Yahy, A., Al-Sumari, A., & Hilal, A. (2021). Preprocessing Arabic text on social media. Heliyon, 7(2), e06191.

[27] El-Affendi, Mohammed Ahmed. "An LVQ connectionist solution to the non-determinacy problem in Arabic morphological analysis: a learning hybrid algorithm." Natural Language Engineering 8.1 (2002): 3-23.

[28] Shrestha, B. B., & Bal, B. K. (2020, December). Named-entity based sentiment analysis of Nepali news media texts. In Proceedings of the 6th workshop on natural language processing techniques for educational applications (pp. 114-120).

[29] Almusaddar, M., 2014. Improving Arabic Light Stemming in Information Retrieval Systems. Thesis MSC Thesis. Computer Engineering Department, Faculty of Engineering, Research and Postgraduate Affairs. Islamic University, Gaza, Palestine

[30] Siswoyo, Siswoyo. "SIMILARITIES AND DIFFERENCES BETWEEN ENGLISH AND ARABIC VERB." Jurnal Smart 2.2 (2016).

[31] Ismail, Samia Ben, Sirine Boukédi, and Kais Haddar. "Transformation system to generate derivational forms of an Arabic verb with HPSG." 2017 International Conference on Engineering & MIS (ICEMIS). IEEE, 2017.

[32] Othman, Mohamed Tahar Ben, Mohammed Abdullah Al-Hagery, and Yahya Muhammad El Hashemi. "Arabic text processing model: Verbs roots and conjugation automation." IEEE Access 8 (2020): 103913-103923.

[33] Mohammed, Rafea. "New Arabic stemming based on Arabic patterns." Iraqi J. Sci. 57.3 (2016): 2324-2330.

[34] Abd Alameer, A. Q. (2017). Finding the similarity between two Arabic texts. Iraqi Journal of Science, 152-162.

[35] Farghaly, A., & Shaalan, K. (2009). Arabic natural language processing: Challenges and solutions. ACM Transactions on Asian Language Information Processing (TALIP), 8(4), 1-22.

[36] Alshalabi, H., Tiun, S., Omar, N., AL-Aswadi, F. N., & Alezabi, K. A. (2021). Arabic light-based stemmer using new rules. Journal of King Saud University-Computer and Information Sciences.

[37] Thalji, N., Hanin, N. A., Al-Hakeem, S., Hani, W. B., & Thalji, Z. (2018). A novel rule-based root extraction algorithm for Arabic language. International Journal of Advanced

Authors' Profiles

Abdulmonem Ahmed A native of Libya received his Bachelor degree in Computer Science in 1993 from the Faculty of Science, University of Benghazi, Libya. From 1996 to 2003, worked for National Commercial Bank- Libya (NCB). In 2008, he obtained a Master degree of Computer Science in Libyan Academy, Tripoli, Libya. In 2019, he enrolled in a PhD program in Computer Engineering, at the Department of Computer Engineering, University of Kastamonu, Turkey.

Aybaba Hançerlioğullari Graduated BSc in physics Engineering department from Hacettepe University/Turkey in 1992. Msc from Gazi University Institute of science Nuclear Research in 1996. PhD from Gazi University institute of science Nuclear Research/Turkey in 2003. He is currently working as Professor at Institute of Science Physics department, Kastamonu University.