Mansoura Journal of Computers and Information Sciences

Arabic characters descriptors for lexicon reduction in Arabic handwriting

Nada Essa
Faculty of computers and information systems I.T dep.
Mansoura University, Egypt
nadaessa2012@gmail.com

Eman El- Daydamony
Faculty of computers and information systems , I.T dep.
Mansoura University, Egypt
eman.8.2000@gmail.com

Ahmed Atwan
Faculty of computers and information systems , I.T dep.
Mansoura University, Egypt
atwan.4@gmail.com

ABSTRACT

This paper introduces an advanced Arabic handwriting recognition technique using lexicon reduction. The lexicon reduction technique stands on extracting the Arabic character shape descriptors. The technique implementation consists of two major stages. The first stage presents a method for extracting the shape descriptor of each character. The second stage suggests Aho-Corasik string searching algorithm for Arabic character recognition. Various stages have been evaluated on the IFN/ENIT database. The results demonstrate the efficiency of the suggested technique.

Keywords

Lexicon reduction, Aho-Corasik algorithm, string searching

1. INTRODUCTION

The Arabic character recognition technique depends on the lexicon because of the confusion and enormous shapes of Arabic writing styles. The lexicon is a particular group of words dependent on the application type. The main problem of the big size lexicon is the long comparison time of the input image with all the words in the lexicon. The lexicon reduction process is removing the lexicon entries that do not match the input image. Lexicon reduction reduces the recognition time. Arabic word lexicon reduction implementation includes two methods similar to a word classifier: holistic and analytical [1]. The holistic method avoids any word segmentation and uses the features of the whole word to recognize it. The analytical method, first splits the input word into characters and then recognize them. In general, lexicon reduction systems differ from classifiers as the classification of the resulting lexicons is not required and the efficient decision is made if a lexicon entry exists in the lexicon entries [2]. The lexicon reduction method depends on extracting the descriptors of the entire Arabic word, sub words or segments. The word shape descriptors define the structure of Arabic word using structural features or the characteristics of Arabic word such as length. Despite the increase of the size of the lexicon, the difficulty of the recognition process and computation complexity raises. This work provides an attempt to reduce and restrict the lexicon entries in addition to the reduction of computational complexity. The lexicon entries are suggested to be the shape descriptors of all the Arabic characters with their different positions. Given an unknown word shape descriptor, a string searching technique is used for determining the positions and the classes of the characters that form the unknown word.

Davoudi et al. [3] presented a lexicon reduction method to compare the input sub-word image with entries of the lexicon, and the highest identical ones were selected. The preferred shape regions of a sub-word are the local regions that differentiate it from other lexicon sub-words. First, a distinction score for each local region was computed, indicating how that region was more fitting. Therefore, these scores were used in a suggested distance measure to modulate the weights of correspondent shape features, and the distinguishing regions were given more weight. A lexicon reduction using global shape dependent on the characterizing loci was used to supplement the local descriptors of sub-word. Davoudi et al. [4] introduced an approach for the reduction of the number of the lexicon entries for printed Farsi sub-words. This method applied the holistic shape with the information of the key character to lower the lexicon size orderly presented using two methods. The first method depended on the description of global shape to construct a pictorial dictionary. The other method based on constituent character information for building a textual dictionary. The reduction step was achieved in two stages. The first stage included the extraction of the characteristic loci features and their comparison with the pictorial dictionary to locate the appropriate sub-words dependent on the similarity of their shapes. Moreover, the lexicon was decreased by detecting the key character of the input image and its comparison with the textual dictionary. Global descriptors and key characters were used to describe the word image. An ideal selection of key characters, depending on the common information for dictionaries of pictorial and textual information, was introduced. The final applicant sub-words are those having the similar key character to the input image.

Chherawala et al. [5] presented a framework for the Arabic word descriptor (AWD) for indexing the shape of an Arabic word and lexicon reduction in documents. This framework provides two stages. Firstly, the structural descriptor (SD) for each connected component (CC) was detected of the input word. The model of bag–of–words was used for defining the CC shape, wherever each visible word represents various local shape structures evolved from the image with filters of various scales and patterns. In addition, the AWD was formed by arranging and equalizing the SDs. The symbolic features of Arabic words were assured, such as diacritics and sub-words and, without segmentation. In the state of lexicon reduction, the AWD was considered as an index to a reference database. Given an image, the reduced lexicon was acquired from the labels of the entries in the indexed database. Parvez et al. [6] introduced a method for lexicon reduction based on descriptors of Arabic segments. An Arabic word was segmented into graphemes, and then a descriptor of the existence of dots in those segments was developed. The segmentation algorithm depend on the details of the structure of Arabic script; and this involved expected Arabic characters segmentations. Moreover, the descriptors of novel canonical segment for lexicon entries were produced. A matching algorithm fitting Arabic handwriting and the resulting descriptors was used for lexicon reduction. Chherawala et al. [7] proposed a lexicon reduction for Arabic documents using a sparse descriptor. The geometrical and topological features of sub-words were obtained from the skeleton image depending on the local density concept. The sparse descriptor was then created as a 3-bins histogram, defining the local density distribution of skeleton pixels. This descriptor was then enlarged to the Arabic word descriptor (AWD), which associated information from all the sub-words and diacritics of an Arabic word.

Chherawala et al. [8] introduced a method of lexicon reduction for handwriting Arabic documents. The shape of the word is denoted by the weighted topological signature vector (W-TSV), and then the graph data were encoded into a low-dimensional vector space. The representatives of three directed acyclic graphs (DAG) were suggested for the shapes of Arabic, dependent on geometrical and topological features. A directed acyclic graph is a graph, without cycles, that has a direction. The search of nearest neighbor in the W-TSV space was used in Lexicon reduction. Wshah et al. [9] proposed a scheme for lexicon size reduction for Arabic word recognition. This scheme evolved the detection of the Piece of Arabic Word (PAWs) and dot descriptors. The position and count of dots and the count of the PAWs were used to remove improbable candidates. Some defined rules were used for the extraction of the dot descriptors and a convolutional neural network was used for verification. A dynamic matching approach and the range of two features were used in the reduction algorithm. Sections of this paper are sorted as follows. The details of the nature of the Arabic character are described in Section 2. Section3 introduces the suggested lexicon reduction technique for Arabic handwriting words. Section 4 discusses the experimentations with the results of the proposed system, and introduces the different situations of wrong recognition. The conclusion and the suggested future work are presented in Sections 5, 6.

2. THE CHARACTERISTICS OF THE ARABIC CHARACTERS.

The descriptions of Arabic character structures contain a big challenge [10]. The Arabic character constructs of two parts. The first part is the main shape of every character and the other part is the marks, and these are a group of diacritics, dots, Hamza, Madda, and Letter Kaf. Various classes are used to categorize the dots. These classes are one or two dots above or under the character, moreover, three dots above the character. The types of diacritics are Vowel, Shadda and Nunation. Vowel diacritics involve three Arabic vowels:Damma, Fatha and Kasra. The Nunation diacritics are considered a duplicate version of their vowels like Dammatan, Fathatan, and Kasratan. Shadda is supposed a doubling diacritic. A set of characters linked vertically or horizontally is called ligatures as shown in Figure1.

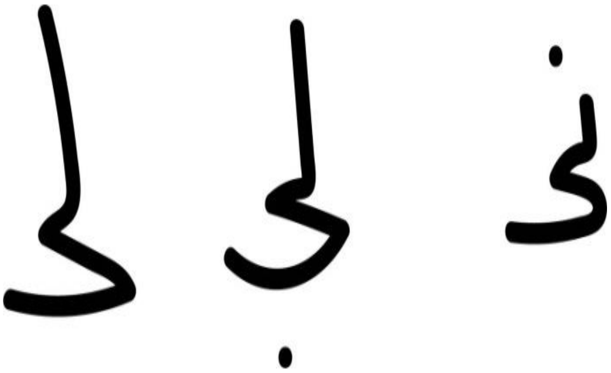
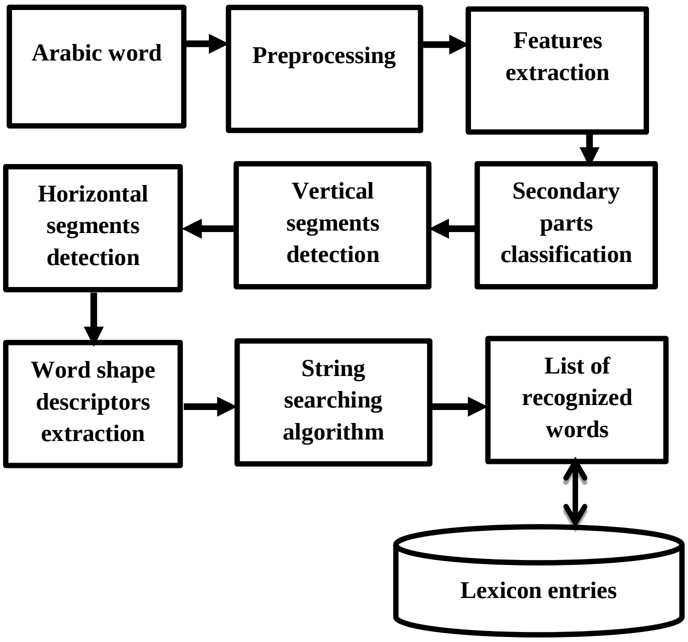


Fig 1: Ligatures shapes.

3. THE PROPOSED LEXICON REDUCTION TECHNIQUE

All previous techniques compare the input word / sub-word image with the lexicon of a defined set of words/ sub-words [1-6]. In this work, the lexicon is a particular set of all Arabic character shape descriptors. The input word descriptor is searched using string searching algorithm to find characters of the input word. The suggested lexicon reduction technique consists of various stages. Firstly the input word is reduced to one pixel wide and the sub-words of each word are extracted. Secondly, for each sub-word, the structural features are detected, the baseline is located, and the secondary parts are detected and classified. Then, the vertical and horizontal suggested segmentation points of each sub-word are located. After that, the word shape descriptors are detected. Finally, the Aho-Corasik string searching algorithm is used for searching for the characters that form the word. These stages are introduced in Figure 2.



Arabic word
Preprocessing
Features extraction
Secondary parts classification
Vertical segments detection
Horizontal segments detection
Word shape descriptors extraction
String searching algorithm
List of recognized words
Lexicon entries

Fig 2: Stages of lexicon reduction technique.

3.1 Preprocessing

The input word is reduced to one pixel wide [11] and the vertical projection is used to detect the sub-words of each word [12]. The baseline of each sub-word is extracted [13]. The primary part and the secondary parts are detected for each sub-word [14]. The secondary parts are distinguished in consonance with their positions to the baseline. Secondary parts are classified using nine different classes as shown in Table 1. Classification principles of secondary parts relied on their description using structural features. When the recognized class is Madda or Shadda, moreover, this distinct secondary part is erased. Figure 3 shows the different secondary parts.

Table 1. Rules for secondary parts Classification.

Class	Description	Rules
2DA	2 Dots Above Baseline	Width>height&& width /height >n, n>1
2DU	2 Dots Under Baseline	
AL	Alif	Height > 2 *width
3D Or HA	Triple Dot Hamza Above Baseline	North cavity or West cavity
HU	Hamza Under Baseline	
MA	Madda Above Baseline	South cavity and North cavity
Shadda	Shadda	South cavity and North cavity or South cavity
DA	Dot Above Baseline	If it does not belong to any of the prior classes
DU	Dot Under Baseline	

Notes: 2Dot Above Baseline= 2DA, 2 Dot Under Baseline= 2DU, Triple Dot = 3D, Hamza Above Baseline = HA, Hamza Under Baseline = HU, Madda Above Baseline = MA, Dot Above Baseline=DA, Dot Under Baseline=DU, Alif= AL.

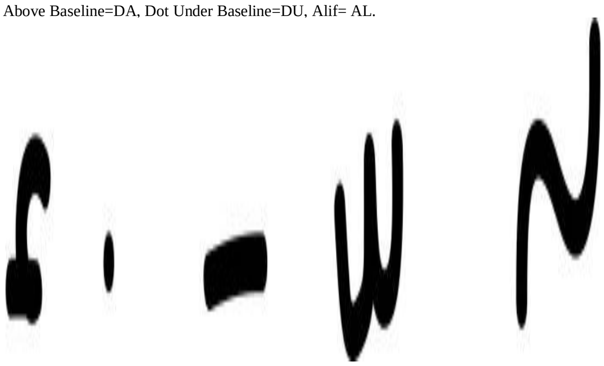


Fig 3: Different types of Arabic secondary parts.

The Arabic characters have various positions. These different positions involve Isolated (I), End (E), Begin (B), and Middle (M), in addition Secondary part class contains two main segments. The first segment involves secondary parts. The second segment accommodates their location as in the following: A indicates Above Baseline, and U represents Under Baseline. Table 2 shows the descriptions of Arabic characters based on the different order of the preceding structural features and secondary parts.

Table 2. Description of Arabic characters based on structural features.

Shape of character	Arabic character shape descriptor			
	B	M	I	E
ا, آ, إ, أ	{AS, HA| HU| MA | NONE}			
ب,ت,ث	{1E, DU|2DA|3D}		{2E, DU|2DA|3D }	
ن	{E|DA}		{ES, W, DA}	
خ , ح ,ج	{W|L, DU|DA |NONE}	{ES|W| L, DU|DA| NONE}	W|L, ES, DU|DA| NONE}	W|L, ES, ES| NONE, DU|DA|NONE}
د، ذ	{E, DA|NONE}			
ر،ز	{E, EU, DA|NONE}			
س، ش	{٣E, 3D|NONE)}		4E, 3D|NONE	
ص,ض	{L, DA| NONE}	{L, ES| NONE, DA |NONE }	{ES, W ,L, DA| NONE}	{ES, W, L, ES| NONE, DA| NONE}
ط, ظ	{AS, L, DA| NONE}	AS, L, ES |NONE, DA| NONE)	AS, L, DA |NONE	{AS, L, ES| NONE, DA| NONE}
ع ,غ	{ES, W|NONE }	{ES|L, W| NONE, ES| NONE}	{ES|L, W, ES}	
ف, ق	{L, W|NONE, DA|2DA}	{L, W| NONE, ES| NONE, DA|2DA}	{ES, L, W|NONE,DA|2DA}	{ES, L, W|NONE, ES|NONE, DA|2DA}
ك	{ES,W}		{ES,AS,HA}	
م	L	{L, ES|NONE}	{ES| NONE, L}	{ES| NONE, L, ES|NONE}
ه	2L		L	L
ة	NONE		L,2DA	
و	L,W			
ؤ	L,W, HA			
ل	AS	ES, AS		
ى	E, 2DU	ES,W,ES,2DU|NONE		
ئ	E, HA	ES,W,ES,HA		
لا,لأ,لإ	NONE	{AS, L, AS, HA|HU|MA }| {W, AS, HA|HU|MA} | {AL, AS, HA|HU|MA}		

Notes: W= West Cavity, ES = East Cavity, AS= Ascender , E = End Point, I = Isolated., B = Begin, E = End, M = Middle.

3.2 Structural Features Extraction

Different structural features are extracted. These various features are Ascender (As), Loop (L) [15], Endpoint (E), and

Cavities [16], moreover, these features are used for the description of Arabic characters as shown in Table 2. Figure 4 shows the different types of structural features.

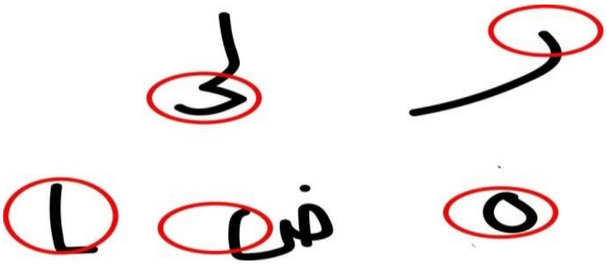


Fig 4: Different structural features.

3.3 Vertical Segments Detection

The aim of this stage is to determine the possible vertical segments of the Arabic word. This algorithm is an improvement of the algorithm introduced by Sari et al. [17]. Firstly, the structural features are determined in each input word image such as end points, ascenders, loops, and cavities. The segmentation points are supposed to be the first columns, where the vertical projection equals zero or one after the top columns of these structural features. Each segment has beginning and ending segmentation points. The supposed vertical segmentation points are as shown in Figure 5.

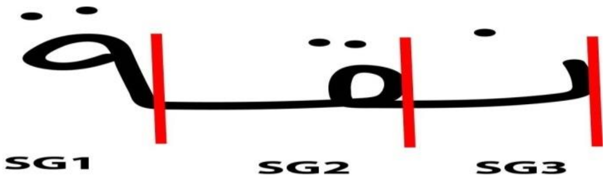


Fig 5: Vertical segmentation points.

There are some exceptions. The first exception when there are three endpoint features, then only the last segmentation point is used to end the segment as shown in Figure 6.(a). If a West cavity feature has an end point under the baseline, then ignore this segmentation point after this feature as shown in Figure 6.(b). If the first extracted feature is the end point or East cavity, then it is ignored as shown in Figure 6.(c).

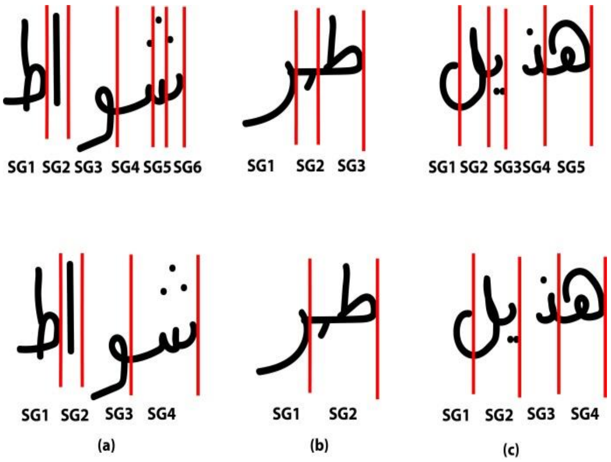


Fig 6:Some exceptions for the detection of the segments.

3.4 Horizontal Segments Detection:

Previously, only the locations for possible vertical segments are detected [6]. This work provides an attempt to locate the possible segmentation points for Arabic ligatures separation. Shapes of Arabic Ligature include two or more connected characters vertically or horizontally. Only these Arabic characters “ج، ح، خ، ى، م” create different styles of Arabic ligatures [18]. Moreover, the structures of these specific characters are described using the loop and West cavity structural features. When the number of loops or West cavities is equal to more than one and there is crossing between them and other detected structural features. Then, the lower raws limits the loop or west cavities features, are added to the supposed horizontal segmentation points for each loop or West cavity feature. Ligatures segmentation points detection is implemented as shown in Figure 7.

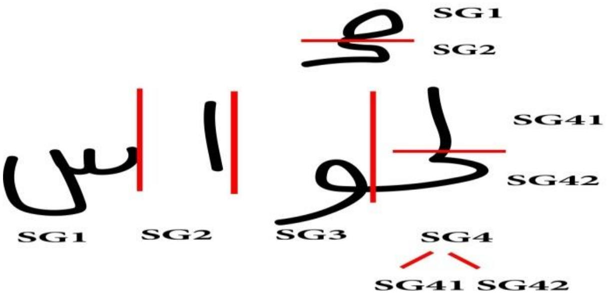


Fig 7: Horizontal segmentation points.

3.5 The Arabic Characters Shape Descriptors

The Arabic characters consist of many graphemes. In linguistics, the grapheme is the smallest unit of the writing system of any language. The character shape descriptors define the characters of the cursive Arabic word. The shape descriptor of every character in the Arabic language is detected based on the structural features and classes of secondary parts in a different order as shown in Table 2.

3.6 Word Shape Descriptor Searching

To construct word descriptors (WD), at first, the structural features are extracted. Secondly, the positions of secondary parts are detected. After that, the class of each secondary body is detmined as dots or diacritic according to the baseline. Finally, the descriptors for Arabic word image are constructed. Assume that a word image P consists of O segments W={SG1, …….., SGo}. Each segment contains M secondary parts and N structural features. The secondary parts (SP) indicate dots or diatricts. The positions of the Structural features (SF) of each segment are detected and sorted from left to right according to their appearance. These structural features are a loop, East Cavity, West cavity, end point, end point under baseline and ascender. The end point is detected in a segment only when there is no any other feature. A descriptor of segment (DSG) consists of two parts {SF, SP}, where SF belongs to a group of structural features {L, W, E, EU, A, ES} and SP belongs to a set of secondary parts {DU, DA, 2DU, 2DA, 3D}. For example, DSG= {L DA} means that the structural feature is a loop and a dot above baseline is the secondary part. The structural features are detected firstly and the secondary parts are detected secondly. Another example is the word" عمران" consists of five segments and its word descriptor is [DSG1, DSG2, DSG3, DSG4, DSG5]. The shape descriptor of each segment is then detected DW = [{ES,ES,DA},{AS},{E},{L},{W,ES}].

3.7 String Searching Algorithm

Other works on lexicon reduction are carried out by matching the descriptors of the unknown word image with the descriptors of the lexicon entries using an algorithm of string matching. In this work, the lexicon entries are considered the descriptors of all Arabic characters with different positions and shapes. Therefore, the number of lexicon entries is reduced. A string searching algorithm is used to search for the segment descriptors in the set of lexicon entries.

3.7.1 Aho–Corasick Algorithm

The algorithm of Aho–Corasickis considered a string searching algorithm [19]. It is considered an algorithm for the dictionary matching that detects elements of a limited group of strings. Given a set of string descriptors of all the Arabic characters with their different positions P. Firstly, the Aho-Corasik automation is constructed by building the tree of a group of strings S as shown in Figure 8. Beginning with the root node only, each string in S is inserted one after the other. Looking up for a string S ={ S_1 ....S_n} starting at the root, and S is considered the word descriptor. The path labeled by elements of S is followed. If the path leads to a node with the identifier. S is considered as a character class in the suggested lexicon entries, otherwise, S is not a character class in the lexicon entries. The output is the character classes of the recognized elements of input Arabic words.

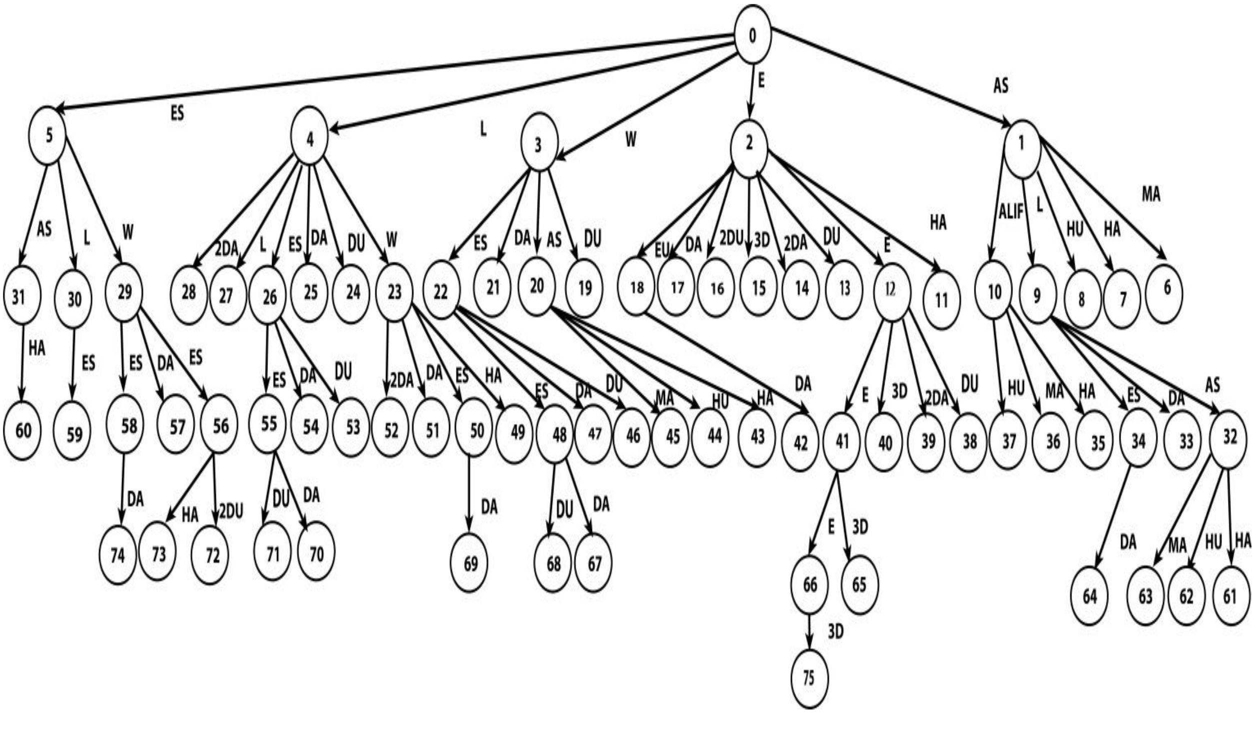


Fig 8: The tree of all Arabic character descriptors.

4. THE EXPERIMENTAL RESULTS

The results of the proposed lexicon reduction technique are presented using the v2.0p1e version of IFN/ENIT database [20]. It includes 32294 samples of 937 Tunisian city names. The sets of the IFN/ENIT database are used for testing the lexicon reduction technique. To calculate the lexicon reducer performance, several standards have been suggested by Madhavanath et al.[21]. Given a set of Q word descriptors WD= {wd_1, wd_2,…………, wdq} and the output is a list of the most identical words X. The lexicon entries are a group of V characters shape descriptor CD= {cd_1, cd_2,…………, CD_V}. So, the lexicon size is minimized to 88 entries. The suggested lexicon reduction technique determines if the unknown word descriptor elements wd_qbelong to candidates of the character descriptors in the lexicon entries . To determine a reduced lexicon wd_q∈ CD. The event that wd_q is formed by the character shape descriptors in the lexicon entries is denoted by a random variable A, where A=1, if wd_q∈ CD, and A=0 else. The range of the reduction is declared by a random variable R, realized as R=(| CD|- |Q |)/|CD|. The main three measures of the efficiency of lexicon reduction are: Accuracy of reduction: α= E(A), Degree of reduction: ρ= E(R), and Reduction Efficacy: η= α.ρ, α,ρ,η € [0,1]. The degree of reduction is inversely associated with the accuracy of reduction. The two measures are collected into one global measure η. Table3 shows The accuracies of reduction for different sets in IFN/ENIT A comparison between the previous lexicon reduction techniques and the proposed technique is given in Table 4.

Table 3. The reduction accuracies for each set in IFN/ENIT.

Set	Reduction accuracy
a	92.97%
b	92.51%
c	90.52%
d	92.32%
e	90.27%

Table 4. Comparison of performance measures with the other recent lexicon reduction methods.

Methods	Performance measures		
	Degree of reduction	Reduction accuracy	Reduction efficacy
AWD [5]	92.1%	90.0%	82.9%
Segments descriptors[6]	90.6%	88.4%	80.1%
Sparse Descriptors [7]	90.1%	90.0%	81.1%
W-TSV [8]	83.6%	90.0%	75.2%
The proposed	92.1%	91.75%	84.50%

The proposed technique errors appeared in two cases. The first case is considered with the errors of the algorithm such as the incorrect classification of secondary parts and the similarity between few characters shapes descriptors as shown in figure 8. For example, the similarities between the shape descriptor of the characters " غ، خ "،"ع، ح" in end positions as shown in Figure 9.(a,b). Another similarity between the shape descriptors of the middle position of the character "ق" and the character "ة" as shown in Figure9.(c). The shape descriptors of the characters "ح، م، ع، و، ص" with the middle positions are identical as shown in Figure9.(d). The similarity between the shape descriptors of different characters " خ، غ، ض", when their position is middle, as shown in Figure9.(e).

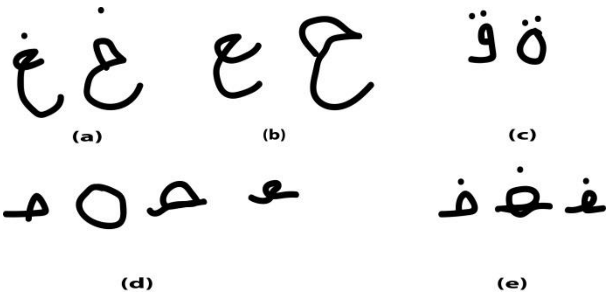


Fig 9: Similarities between characters shape descriptors.

The other case appears with bad word writing, the loss of the main shape of the characters as shown in Figure 10.

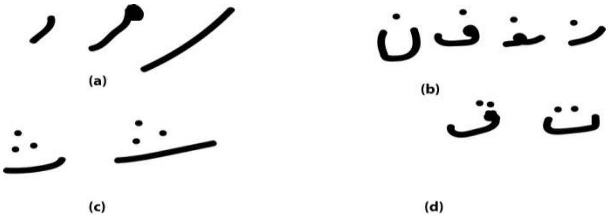


Fig 10: Characters shape loss.

Another case emerges when there is an incorrect gathering of secondary parts as shown in Figure11

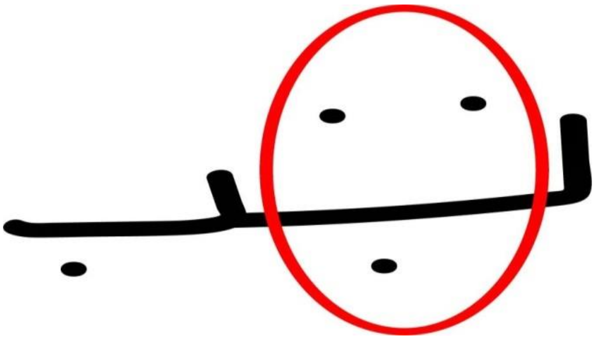


Fig 11: Incorrect grouping of secondary parts.

5. THE CONCLUSION

This paper presents a new lexicon reduction technique for Arabic word using the Arabic characters shape descriptors. Firstly, the shape descriptor of Arabic word is determined. Secondly, the Aho-Corasik string searching algorithm is used to search for the characters shape descriptors of the Arabic word in the lexicon entries. The proposed technique is tested using IFN/ENIT database and the final results are promising and indicate enhancement of the accuracy of reduction

6. THE FUTURE WORK

We will focus in the future work on trying to overcome the different challenges which faced the proposed lexicon reduction algorithm. Also, different string searching algorithms would be used to enrich the shape descriptors of the characters with different detailed features in their description.

7. REFERENCES

[1] Mozaffari, S., Faez, K., Märgner, V, El-Abed, H. 2008. Lexicon reduction using dots for off-line Farsi/Arabic handwritten word recognition. Pattern Recognition Letters VOL.29, NO.6, 724-734.

[2] Mozaffari, S., Faez, K., Maergner, V., El Abed, H. 2008. Two-stage lexicon reduction for offline Arabic handwritten word recognition. International Journal of Pattern Recognition and Artificial Intelligence VOL.22, No.07, 1323-1341.

[3] Davoudi, H., Cheriet, M., Kabir, E. 2016. Lexicon reduction of handwritten Arabic subwords based on the prominent shape regions. International Journal on Document Analysis and Recognition (IJDAR). VOL. 19, No.2, 139-153.

[4] Davoudi, H., and Kabir, E. 2014. Lexicon reduction for printed Farsi subwords using pictorial and textual dictionaries. International Journal on Document Analysis and Recognition (IJDAR). VOL. 17, No.4, 359-374.

[5] Chherawala, Y., and Cheriet, M. 2014. Arabic word descriptor for handwritten word indexing and lexicon reduction. Pattern Recognition VOL.47, NO.10, 3477-3486.

[6] Parvez, M. T., and Mahmoud, S. A. 2013. Lexicon Reduction Using Segment Descriptors for Arabic Handwriting Recognition. 12th International Conference on Document Analysis and Recognition (ICDAR), IEEE.

[7] Chherawala, Y., Wisnovsky, R., Cheriet, M. Sparse descriptor for lexicon reduction in handwritten Arabic documents. 21st International Conference on Pattern Recognition (ICPR), IEEE.

[8] Chherawala, Y., and Cheriet, M. 2012 .W-TSV: Weighted topological signature vector for lexicon reduction in handwritten Arabic documents. Pattern Recognition VOL.45, NO.9, 3277-3287.

[9] Wshah, S., Govindaraju, V., Cheng, Y., Li, H. 2010. A novel lexicon reduction method for Arabic handwriting recognition. 20th International Conference on Pattern Recognition (ICPR), IEEE.

[10] Habash, N. Y. 2010. Introduction to Arabic natural language processing. Synthesis Lectures on Human Language Technologies.

[11] Lam, L., Lee, S. W., Suen, C. Y. 1992. Thinning methodologies-a comprehensive survey. IEEE Transactions on pattern analysis and machine intelligence VOL.14, NO.9 , 869-885.

[12] Maliki, M., Jassim, S. 2012. Arabic handwritten: pre-processing and segmentation. SPIE Defense, Security, and Sensing.

[13] Abu-Ain, T., Abdullah, S. N. H. S., Bataineh, B., Abu-Ain, W., Omar, K. 2013. Text normalization framework for handwritten cursive languages by detection and straightness the writing baseline. Procedia Technology VOL.11 .666-671.

[14] Parvez, M. T., and Mahmoud, S. A. 2013. Arabic handwriting recognition using structural and syntactic pattern attributes. Pattern Recognition VOL.46, NO.1, 141-154.

[15] Touj, S. M., Amara, N. E. B., and Amiri, H. 2005. Arabic handwritten words recognition based on a planar hidden Markov model. International Arab Journal of Information Technology. VOL.2, NO.4, 318-325.

[16] Dougherty, E. R. 1994. Digital image processing methods. Marcel Dekker, Inc.

[17] Sari, T., Souici, L., and Sellami, M. 2002. Off-line handwritten Arabic character segmentation algorithm: ACSA . Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition, IEEE.

[18] Sari, T., and Sellami, M. 2007. Overview of Some Algorithms of Off-Line Arabic Handwriting Segmentation. International Arab Journal of Information Technology. VOL.4, NO.4 ,289-300..

[19] Navarro, G., and Raffinot, M. 2002. Flexible pattern matching in strings: practical on-line search algorithms for texts and biological sequences. Cambridge University Press.

[20] Pechwitz, M., Maddouri, S. S., Märgner, V., Ellouze, N., Amiri, H. 2002, IFN/ENIT-database of handwritten Arabic words. Colloque International Francophone sur l'Ecrit et le Document (CIFED) VOL.2. 2002.

[21] Madhvanath, S., Krpasundar, V., Govindaraju, V. 2001. Syntactic methodology of pruning large lexicons in cursive script recognition. Pattern Recognition VOL.34, NO.1, 37-46.