To recognize the manuscript texts of Arabic letters in ancient Uzbek script

Iskandarova Sayyora Nurmamatovna

ICT Center under the Tashkent University of Information Technology named after
Muhammad al-Kharazmiy, Tashkent City, Uzbekistan

E-mail address: sayyora5@mail.ru

ABSTRACT

This article describes the Hemming method of the neural model for automatic identification of Arabic texts on the computer. The main problem of recognizing manuscript mantles in Arabic is that of the elements that they have created. Usually, the text is divided into rows, and then separated by separate words. The development of the Arabic language signifies a great deal of controversy over Arabic language. Hemming is based on the neuronal model and the description of the software product.

Keywords: Hemming, method, neural model, Arabic text, software product, recognize, neural scheme, biologic neural

1. INTRODUCTION

The Arabic writing, which has been a part of Arabic history in the history, has played an important role in the cultural life of Iranians and Turks for centuries, has passed through the Arab tribes of the Arabs before the adoption of the Islamic religion, and the Babachids derived it from the Arabs and the Orthomians from the Phoenicians. Both the Naboths and the Arabs, like the Arabs, were the peoples of the Somali language, their 22-letter consonant (the expression of consonants) corresponded to the Arabic language.

Later on, before the spread of the Islamic religion, Arabs add six letters representing specific Arabic sounds. When the peoples of Central Asia accepted Arabic letters, there are 28 letters in it, most of the letters (Arabic letters) are Arabic to serve the expression of the special voices of his language, and the indigenous peoples can not pronounce those sounds exactly as they were used in the Arabic words. Even though the 28-year-old Arabs have been accustomed to reading Arabic texts, they have difficulty expressing their ideas in their language. This will make it necessary to insert some additional inputs (p چ ژ غ). Thus, the 32-letter inscription was formed.

This article has served for over a thousand years in Central Asia, and it is well-known that rich written cultural artifacts have been created. It is also true that after 1917, this writing was reformed, and the transition to a new record of Latin and Slavic inscriptions led us to the remnants of the centuries-old ancient record and the precious Today, the attitude to the cultural heritage and the study of the ancient Uzbek record, which is the key to the inheritance of this inheritance, has controversy in the issue of naming the record, which has passed through Arabs to our nation's historical record. Some think that this is a direct "Arab script," and they also call it "Arabic - Uzbek inscription", written by Arabs.

More than a thousand years have passed since this historic event. First, the spiritual man of the text had changed. Now the article is not only the Arabs but also the Iranian and Turkic peoples. Secondly, the writing itself has changed, letters have not been added to the Arabic letters, its graphics capacities have expanded and improved.

In Arabic, the inscription was a consonant (representing consonant voices). In the Turkic peoples, especially in the history of the Uzbek people, it has been transformed into a writer, which is based on ancient Turkic writing traditions, which has changed its basis. According to these and similar reasons, we can not call the ancient record used in the history of our people as the "Arabic script."

It is desirable to name an Arabic-language record - a Persian form, a Turkic-language inscription - "Turkic writing." However, the "Old Turkic writing" may also include the ancient Turkic inscriptions, such as the Urhun-Enasoy inscription, the Uyghur script. In order to differentiate between the concepts of "Old Turkic language", "Old Turkic writing", they used the term "old-fashioned Uzbek language" in relation to the history of Uzbek language, so our great ancestors, such as Lutfi, Alisher Navoi, Bobur, We call the ancient Turkic record on the basis of the Arabic alphabet, the oldest Uzbek inscription, to the next generation. The same can be said about the history of other peoples who made great spiritual treasures in the old writing.

When alphabet comes to the beginning and end of the Arabic words, alif does not mean pure sound, just like the words 'aza', 'ash-i'tâ' means a sounds like a hamster (ء) with a barrier. Like in our old writing, letters are written together in the old Uzbek script. Attachment or additionally, the writing result may vary depending on the place of the word: the word comes at the beginning of the word - the same as the word - the second-word third - the third kind. There are four different types.

Here we first get acquainted with a table showing the number of letters and forms of the old Uzbek inscription, and on this basis, we look at the simplest and most complex aspects of writing.

Name of letter	View of letter	Subscription of letter			
		at the end	at the middle	at the beginning	alone
Alif	O,a	ـا	ـا	ا	ا
Be	B	ـب	ـبـ	بـ	ب
Pe	P	ـپ	ـپـ	پـ	پ
Te	T	ـت	ـتـ	تـ	ت
Se	S	ـث	ـثـ	ثـ	ث
Jim	J	ـج	ـجـ	جـ	ج
Chim	Ch	ـچ	ـچـ	چـ	چ
Ho-ye hutti	H	ـح	ـحـ	حـ	ح
Xe	X	ـخ	ـخـ	خـ	خ
Dol	D	ـد	ـد	د	د
Zol	Z	ـذ	ـذ	ذ	ذ
Re	R	ـر	ـر	ر	ر
Ze	Z	ـز	ـز	ز	ز
Же	ж	ـژ	ـژ	ژ	ژ
Sin	S	ـس	ـسـ	سـ	س
Shin	Sh	ـش	ـشـ	شـ	ش
Sod	S	ـص	ـصـ	صـ	ص
Zod	Z	ـض	ـضـ	ضـ	ض
To	T	ـط	ـطـ	طـ	ط
Zo	Z	ـظ	ـظـ	ظـ	ظ
'ayn	-	ـع	ـعـ	عـ	ع
G'ayn	G'	ـغ	ـغـ	غـ	غ

2. METHODOLOGY

Synthetic neurons imitate the biological neuronal properties of the first approach. Every artificial neuron contains a set of signals that other neurons can exit. The synaptic force multiplies each incoming signal, and the sum of them determines neuron activity. The w1 multiplies each signal, w2, ... , wn corresponding to it, and the collector with the S falls into the block. Each weight corresponds to the "power" of a single biological synapse. The W vector determines the weight set. The collecting block corresponding to the body of the biological helmet collects algebraic accumulated inflated values into the appropriate mass and forms the neuronal output. The Net sets this amount (Figure 1).

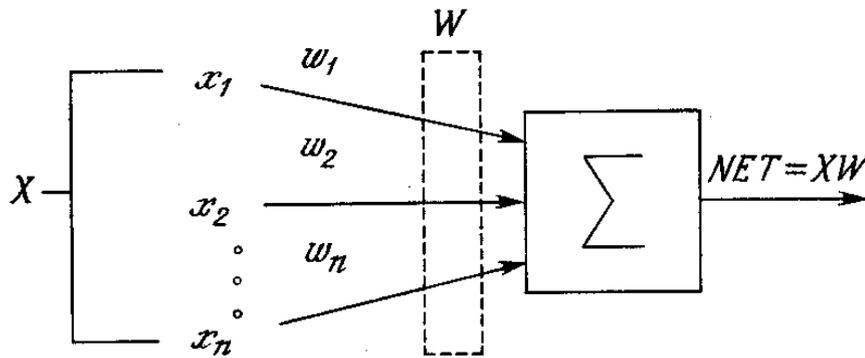The vectors in the above view are as follows:

$$NET = XW.$$

Figure 1. Artificial neural scheme

Activation Functions. In the next step, the NET signal is usually generated by the F activation function and generates the output signal of the neuron. The activation function can be a usual linear function

$$OUT = K\ (NET),$$

where: K – is the constant defined as the following boundary value function

$$OUT = 1, if\ NET > T,$$

$$OUT = 0\ for\ other\ purposes,$$

where: T – is a constant boundary constant. The activation function can be a function that fully describes the nonlinear conductivity of the biological neurons and can provide broader opportunities for the neural network.
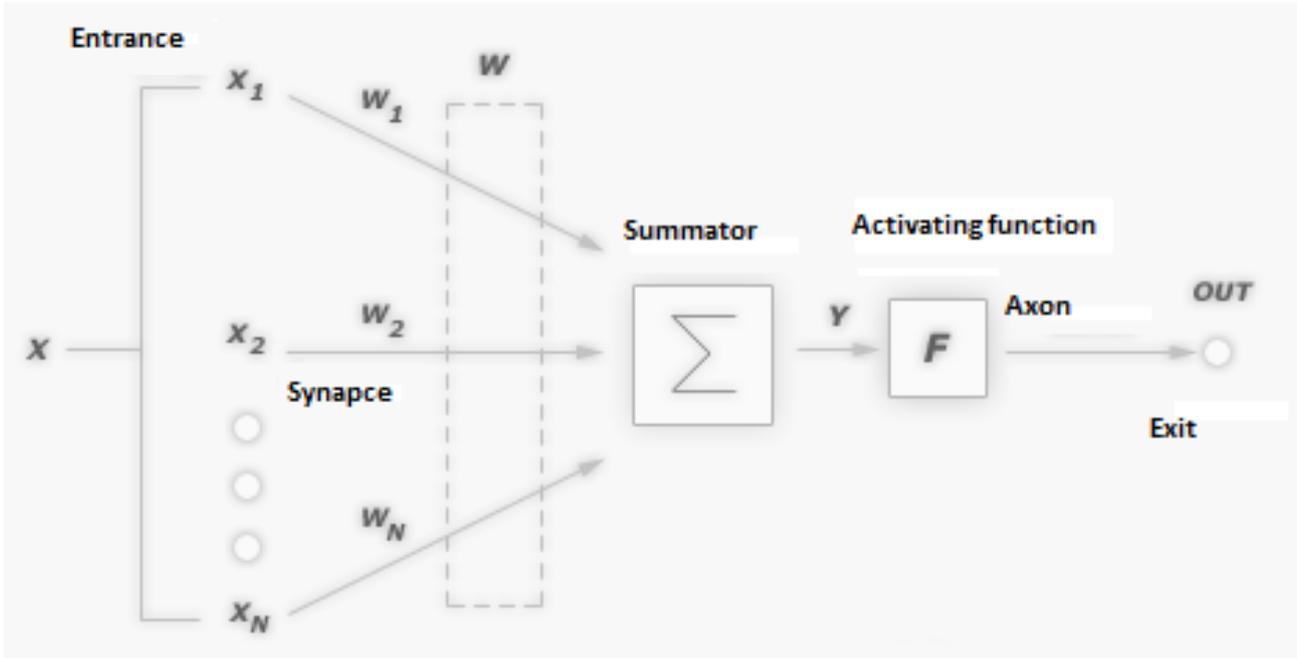
Figure 2. Synthetic neuron with the activation function

The block defined by F in Figure 2 accepts the NET signal and outputs the OUT signal. If the F block encircles the range of change in the NET size, i.e. at any value of the net value, OUT refers to an intermediate range, then the function F is called "compression." In most cases, the logical or "sigmoidal" (S-shaped) function, as shown in Figure 3, is used as a "bulk" function. Mathematical view of this function is F (x) q 1 / (1 + ye-x). And so,

$$OUT = \frac{1}{1+e^{-NET}}.$$

The function of activation from the point of view of similarity with electronic systems can be considered as a non-linear amplifier property of artificial neurons. The amplifier override is calculated as the ratio of OUT size to the relative size of the net value generated by it. When the smaller signal processing problem is solved in areas that are at the center of high-powered logic functionality, low-power amplifiers in positive and negative the signals are enhanced and vice versa, the larger the signals will be reused for reducing the effects. Thus, the neuron operates with a large amplifier in a wide range of unwanted signals, that is, low.
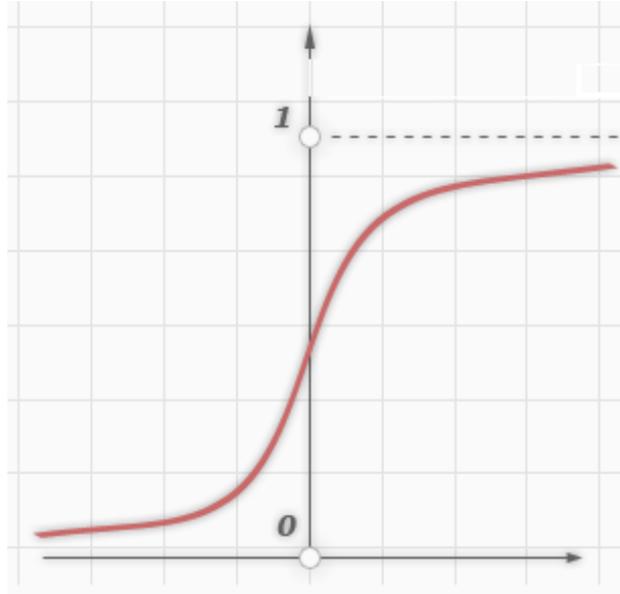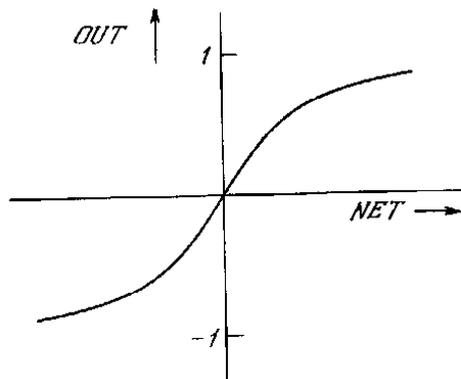
$$OUT = \frac{1}{1+e^{-NET}} = F(NET).$$

Figure 3. Sigmoidal logistics function

Another widely used activation function is hyperbolic tangent. It is similar to the logistic function and is used as a mathematical model for biological activation of the neural network. In the form of an artificial neural network activation function it is written:

$$OUT = th(x).$$

Picture 4. Hyperbolic tangent function

The hyperbolic tangent function is a S-shaped function like logistic functions, but it's symmetric relative to the coordinate and NET is set to zero at OUT output at point 0 (Figure 4). In contrast to the logistic function, hyperbolic tangents accept values in different gaps, and this can be used for some tracks. The simplest artificial neural model denies some of the properties of biological neurons. For example, it ignores the time stops that affect the system dynamics.

Incoming signals immediately generate outgoing signals. And the effect of a very important frequency modulation function or synchronous function of a biological neuron is not taken into account, although some of these features are considered as decisive. Despite these limitations, neurons generated by such neurons show many properties that satisfy the biological system.

2. 1. The concept of Hemingway's second line

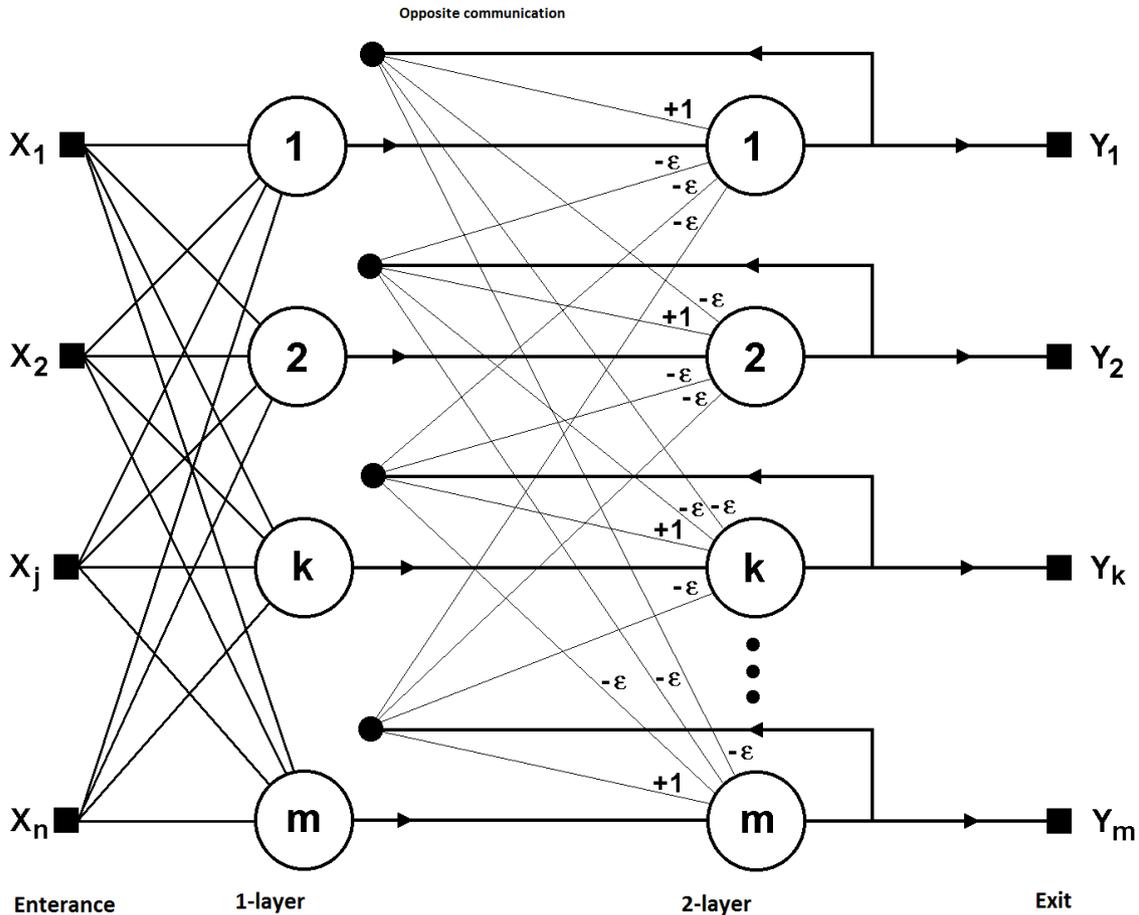The structure of the Hemmergene neuronal network is as follows:

Figure 5. Structural scheme of Hemming network.

If the network generates the exact number of the form, then the associative memory will use the Hemming network sufficiently (Figure 5).

The network consists of two layers. The first and second layers consist of neurons, where the number of forms. The neurons in the first layer consist of synapses that are associated with n networking. The neurons in the second layer are linked to the synapses that are in contact. The only positive, positive feedback is coupled with its neurons.

The purpose of a network work is to find the cache between the images being tested by all the images. The distance from the Hemming is the difference in bits of these two binary

vectors. The network selects the minimum distance between the samples and the missing images, and as a result, a neuron that fits the image becomes active.

Assessing the weighted average coefficients for the first layer and evaluating the activating function is carried out as follows

$$w_{ik} = \frac{x_i^k}{2}, \ i=0...n\text{-}1, \ k=0...m\text{-}1$$

$$T_k = n \ / \ 2, \ k = 0...m\text{-}1$$

where: x_i^k –k- type i-element.

Synapse, stiff weight coefficients, receive a value of $0 < \varepsilon < 1/m$ it accepts the value. The neural synapse associated with its axis is +1.

Xemming tarmoqining bajaradigan ishni algoritmi qu'yidagicha:

1) The node that has been calculated on the first layer of the network and its value $\mathbf{X} = \{x_i : i=0...n\text{-}1\}$, the vector is inserted (indicates the upper layer layer number):

$$j=0.. \ y_j^{(1)} = s_j^{(1)} = \sum_{i=0}^{n-1} w_{ij} x_i + T_j \ .m\text{-}1$$

Then the axis of the second layer is based on the obtained values:

$$y_j^{(2)} = y_j^{(1)}, \ j = 0...m\text{-}1$$

2) The new state of the second layer is:

$$s_j^{(2)}(p+1) = y_j(p) - \varepsilon \sum_{k=0}^{m-1} y_k^{(2)}(p), k \neq j, j = 0...m-1$$

and the value of its axons:

$$y_j^{(2)}(p+1) = f\left[s_j^{(2)}(p+1)\right], \ j = 0...m-1$$

The value of the activator f is [-1; 1], where F is sufficiently large and the value of an arbitrary argument does not exceed.

3) In the last iterative it is checked whether the value of the second layer varies, or if yes, then step 2, otherwise the cycle is stopped.

The evaluation of the algorithm shows that the first layer is of paramount importance: in the first step we use the values of the coefficients of the weight, so the first layer is removed from the network as it is indicated. The Xmmming Network programming model is based on the selection of special classes.

2. 2. The working principle of the Arabic Alphabet Recognition Program

Due to the specific language of the Arabic language, there are still many problems. That's why we recommend using a multi-dimensional optimization method of xamming. Depending on the above problems, there is a need to develop new and effective utility programs. Therefore, we will consider the process of defining the Arabic alphabet program and its recognition program in c # programming environment.
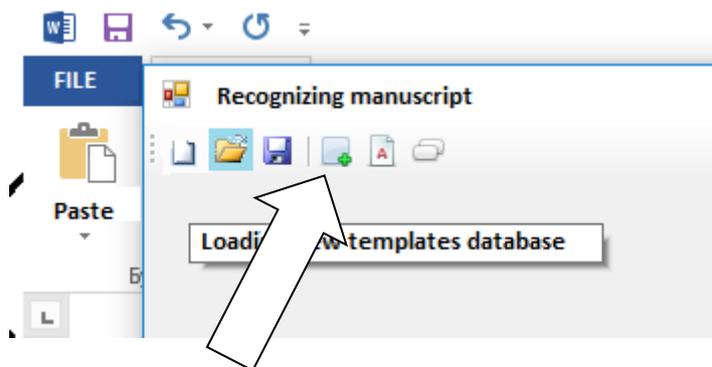
The main window of the program
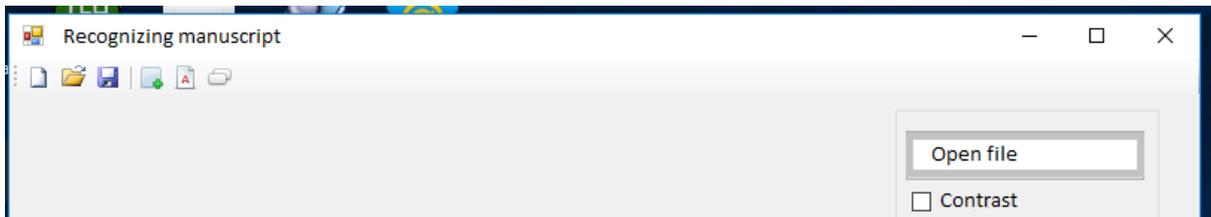
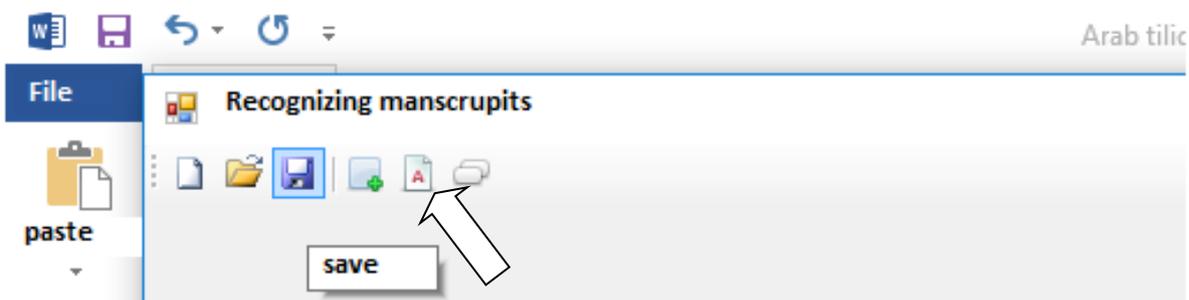By opening the file, the image you want is taken as an image.

The text to be read should be selected from the template and select the text that should be recognized.
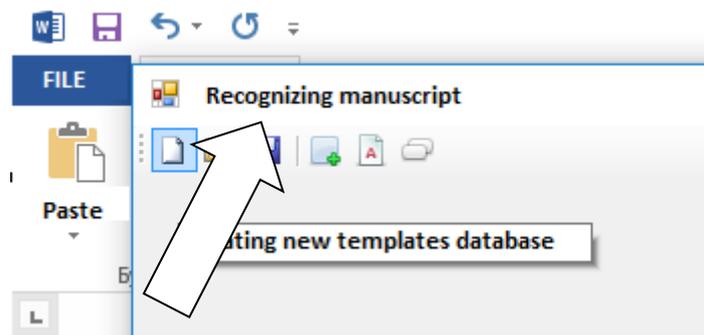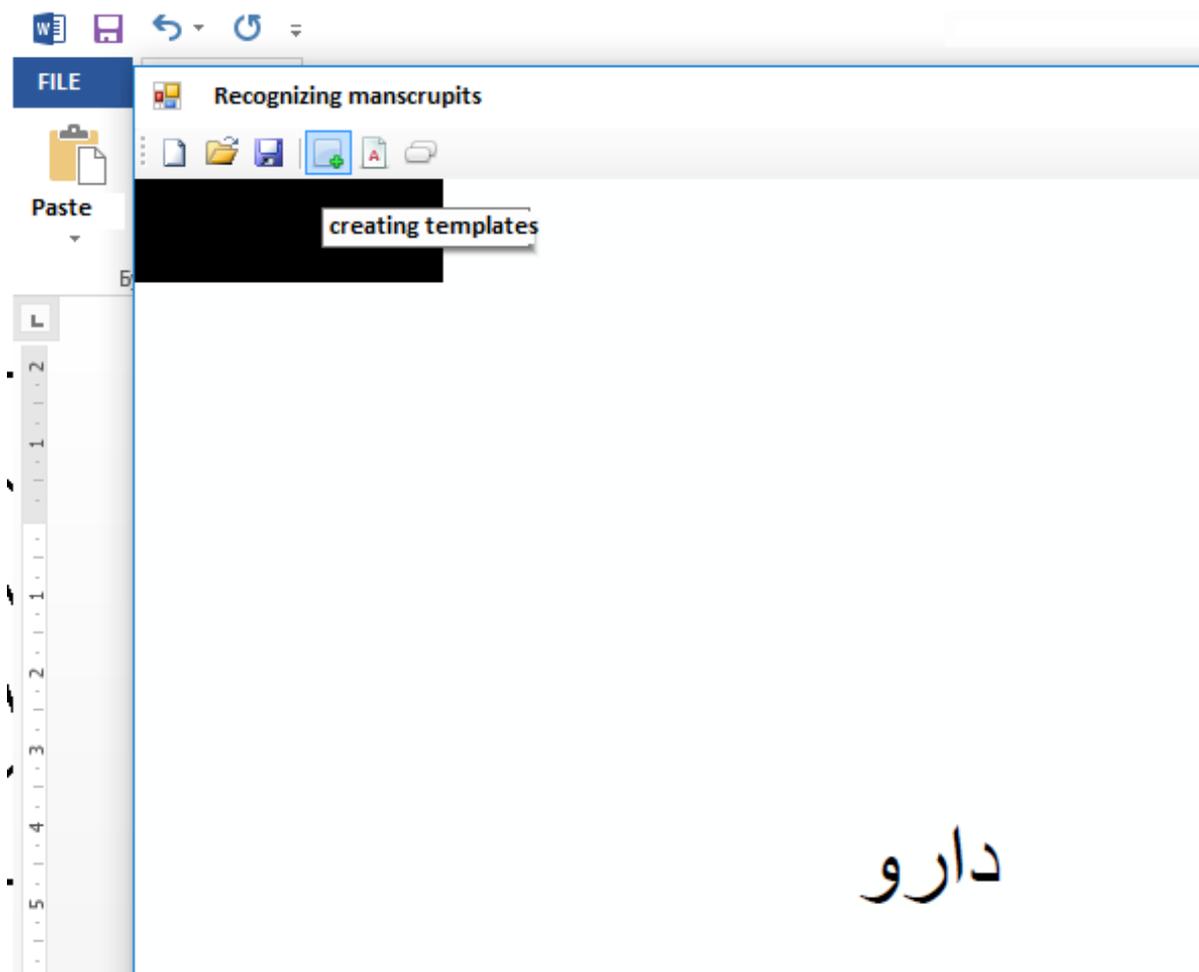

Creating new templates allows you to read letters. The new template database is automatically created.


In the case of a pre-existing database, you can download and configure it.

New templates created by saving are saved.

Before choosing to create a template, open the text you want to copy to the work area and then read the template. Each letter is separated from the open space in the space and the new template is read out to the database. When designing the image, the database is selected and made familiar to it.
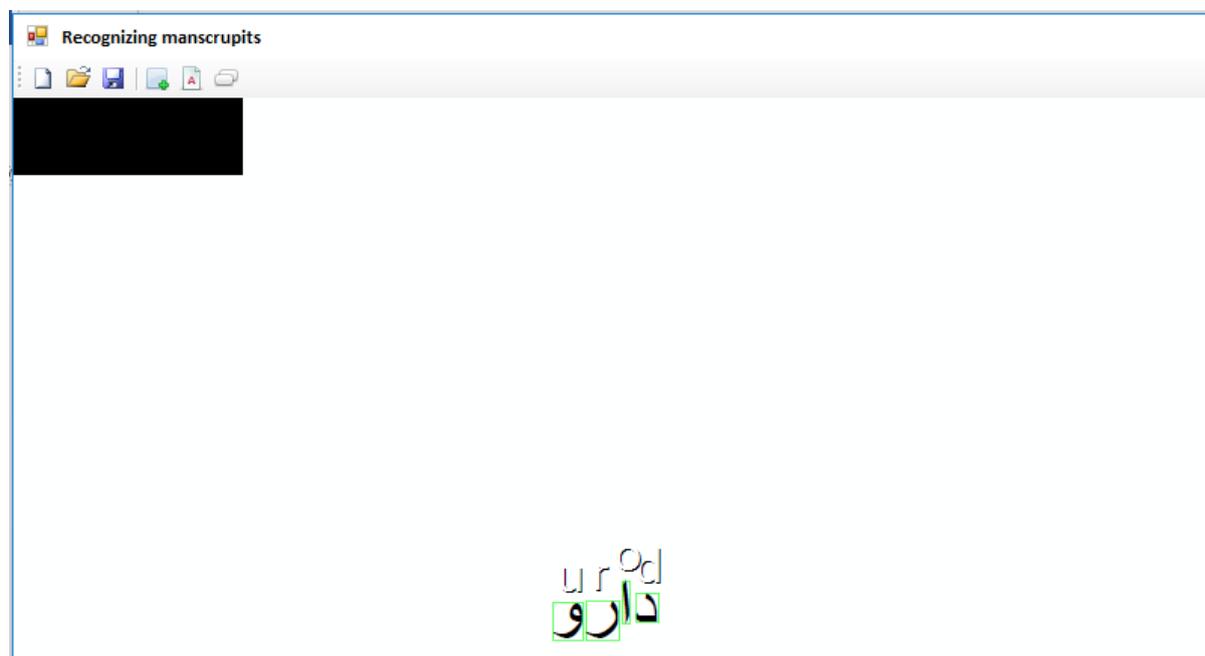
Because of the difference in Arabic letters and Latin graphics, the old Uzbek writing that we have written is the result. This is done on the basis of the diagnosis. The Arabic position of the dot points changes the names of the characters and thus makes it difficult to identify the program and performs 80% recognition.

3. CONCLUSION

When you read the various templates and font templates in the article, you get the familiar Arabic script. It is desirable to use these methods with morphological features of the Arabic language. There are several ways in which the Arabic language is different. This article is not intended to discuss the morphology of the Arabic language, but rather the form of manuscript, which is how computer alogorites are understood. In the original writing of the old Arabic inscriptions, the different types and font-size templates were studied on the basis of the Xemming Neural model. It was done using the opencv library to organize the images on a dotted basis and create graphics, and store them in the image.

References

[1]   S. Young, al., the HTK Book V3.4 Cambridge University Press, Cambridge UK, 2006
       M. T. Parvez, A. M. Sabri, Arabic handwriting recognition using structural and
       syntactic pattern attributes. *Pattern Recognition* 46 (2013) 141-154.

[2]   M. Azmi and R. S. Almajed. A survey of automatic Arabic diacritization techniques.
       Natural Lang. Eng. 21(3) 2013) 477–495.

[3] S. Ananthakrishnan, S. Narayanan, and S. Bangalore. Automatic diacritization of arabic transcripts for automatic speech recognition. Proc. 4th Int. Conf. Natural Lang. Process. Kanpur, India, 2005, pp. 1–8.

[4] M. Elshafie, H. Almuhtaseb, and M. Alghamdi. Techniques for high quality Arabic speech synthesis. *Inf. Sci.* vol. 140, pp. 255–267, Feb. 2002.

[5] Rebai and Y. BenAyed. Text-to-speech synthesis system with Arabic diacritic recognition system. *Comput. Speech Lang.* vol. 34, no. 1, pp. 43–60, 2015.

[6] Zouaghi, L. Merhbene, and M. Zrigui, ''Combination of information retrieval methods with LESK algorithm for Arabic word sense disambiguation. *Artif. Intell. Rev.* vol. 38, no. 4, pp. 257–269, 2017.

[7] Farghaly and K. Shaalan. Arabic natural language processing: Challenges and solutions. *ACM Trans. Asian Lang. Inf. Process* vol. 8, no. 4, pp. 1–22, 2009

[8] M. Maamouri and A. Bies. The Penn Arabic Treebank. Proc. Arabic Comput. Linguistics, Stanford, CA, USA, 2010, pp. 1–42.

[9] O. Bahanshal and H. S. Al-Khalifa. A first approach to the evaluation of Arabic diacritization systems. Proc. 7th Int. Conf. Digit. Inf. Manage. (ICDIM), Macau, China, Aug. 2012, pp. 155–158

[10] Y. A. El-Imam. Phonetization of Arabic: rules and algorithms. *Comput. Speech Lang.* vol. 18, no. 4, pp. 339–373, 2003.

[11] M. Attia. A large-scale computational processor of the Arabic morphology, and applications. M.S thesis, Faculty Eng., Cairo Univ., Giza, Egypt, 2000.

[12] M. Attia. Handling Arabic morphological and syntactic ambiguity within the LFG framework with a view to machine translation. Ph.D. dissertation, School Lang., Linguistics Culture, Univ. Manchester, Manchester, U.K., 2008.

[13] M. Badrashiny. Automatic diacritizer for Arabic texts. M.S. thesis, Faculty Eng., Cairo Univ., Giza, Egypt, 2009.

[14] G. Abandah, A. Graves, B. Al-Shagoor, A. Arabiyat, F. Jamour, and M. Al-Taee, Automatic diacritization of Arabic text using recurrent neural networks. Proc. 13th Int. Conf. Document Anal. Recognit. Nancy, France, vol. 18, 2015, pp. 183–197.

[15] Y. Gal. An HMM approach to vowel restoration in Arabic and Hebrew. Proc. ACL Workshop Comput. Approaches Semitic Lang., 2002, pp. 27–33.

[16] A. El-Harby, M. A. El-Shehawey, and R. El-Barogy. A statistical approach for Qur'an vowel restoration. *J. Artif. Intell. Mach. Learn.* vol. 8, no. 3, pp. 9–16, 2008.

[17] M. S. Khorsheed. A HMM-based system to diacritize Arabic text. *J. Softw. Eng. Appl.* vol. 5, pp. 124–127, Dec. 2017.

[18] M. Elshafei, H. Al-Muhtaseb, and M. Alghamdi. Statistical methods for automatic diacritization of Arabic text. Proc. 18th Nat. Comput. Conf., Riyadh, Saudi Arbia, 2006, pp. 301–306.

[19] M. Elshafei, H. Al-Muhtaseb, and M. Alghamdi. Machine generation of Arabic diacritical marks. Proc. Int. Conf. Mach. Learn., Models, Technol. Appl., Las Vegas, NV, USA, 2006, pp. 128–133.

[20] R. Nelken and S. M. Shieber. Arabic diacritization using weighted finitestate transducers. Proc. ACL Workshop Comput. Approaches Semitic Lang., 2005, pp. 79–86.

[21] Y. Hifny. Higher order n-gram language models for Arabic diacritics restoration. Proc. 12th Conf. Lang. Eng. (ESOLEC), Cairo, Egypt, 2012, pp. 1–5.

[22] Zitouni, J. Sorensen, and R. Sarikaya. Maximum entropy based restoration of Arabic diacritics. Proc. 21st Int. Conf. Comput. Linguistics 44th Annu. Meeting ACL, Sydney, NSW, Australia, 2006, pp. 577–584.

[23] Zitouni and R. Sarikaya. Arabic diacritic restoration approach based on maximum entropy models. *Comput. Speech Lang.* vol. 23, no. 3, pp. 257–276, 2009.

[24] M. Alghamdi et al., Automatic Arabic text diacritizer. King Abdulaziz City Sci. Technol., Riyadh, Saudi Arabia, Tech. Rep. CI.25.02, 2006.

[25] P. De Doncker. A volume/surface potential formulation of the method of moments applied to electromagnetic scattering. *Eng. Anal. Boundary Elements,* vol. 27, no. 4, pp. 325–331, 2003.

[26] F. Vico, L. Greengard, M. Ferrando, and Z. Gimbutas. The decoupled potential integral equation for time-harmonic electromagnetic scattering. *Commun. Pure Appl. Math.* vol. 69, no. 4, pp. 771–812, Apr. 2016.

[27] W. C. Chew. Vector potential electromagnetics with generalized gauge for inhomogeneous media: Formulation (invited paper). *Prog. Electromagn. Res.* vol. 149, pp. 69–84, 2014.

[28] Q. S. Liu, S. Sun, and W. C. Chew. A vector potential integral equation method for electromagnetic scattering. Proc. Int. Rev. Prog. Appl. Comput. Electromagn. (ACES), Mar. 2015, pp. 1–2.

[29] D. Jackson, Classical Electrodynamics. San Rafael, CA, USA: Morgan & Claypool, 2008.

[30] O. Ergül and L. Gürel, The Multilevel Fast Multipole Algorithm (MLFMA) for Solving Large-Scale Computational Electromagnetics Problems. New York, NY, USA: Wiley, 2014.

[31] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Nemurica* vol. 14, pp. 1–137, Apr. 2005.

[32] C. Keller, N. I. M. Gould, and A. J. Wathen. Constraint preconditioning for indefinite linear systems. *SIAM J. Matrix Anal. Appl.* vol. 21, no. 4, pp. 1300–1317, 2000.

[33] R. S. Varga, Geršgorin and His Circles. Berlin, Germany: Springer-Verlag, 2004.

[34] Anne-Laure Bianne-Bernard, Fares Menasri, Laurence Likforman-sulem, Chafic Mokbel, Christopher Kermorvant (2012) Variable length and context-dependent HMM

letter form models for Arabic handwritten word recognition. In Document Recognition and Retrieval Conference (DRR)

[35] AL-Shatnawi, M. Atallah, AL-Salaimeh, Safwan, AL-Zawaideh, Farah Hanna, Omar, Khairuddin, Offline arabic text recognition an overview. *World Comput. Sci. Inform. Technol. J.* 1 (5), 184–192, 2011.

[36] H. Alkhateeb, O. Pauplin, J. Ren, J. Jiang, Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition. *Knowledge-Based Systems* 24 (2011), pp. 680-688.

[37] W. Chew, M.-S. Tong, and B. Hu, Integral Equation Methods for Electromagnetic and Elastic Waves. San Rafael, CA, USA: Morgan & Claypool, 2008.

[38] F. P. Andriulli, A. Tabacco, and G. Vecchi, Solving the EFIE at low frequencies with a conditioning that grows only logarithmically with the number of unknowns. *Antennas Propag.* vol. 58, no. 5, pp. 1614–1624, May 2010.

[39] Lawgali. A Survey on Arabic Character Recognition. *International Journal of Signal Processing, Image Processing and Pattern Recognition* Vol. 8, No. 2 (2015), pp. 401-426.