

# Document Recognition Method Based on Convolutional Neural Network Invariant to 180 Degree Rotation Angle\*

E. I. Andreeva<sup>1</sup>, V. V. Arlazarov<sup>1,II,III</sup>, A. V. Gayer<sup>II</sup>, E. P. Dorokhov<sup>1</sup>, A.V. Sheshkus<sup>1,II,III</sup>, O.A. Slavin<sup>I, II,III</sup>

<sup>1</sup> Smart Engines Service LLC, Moscow, Russia,

<sup>II</sup> Federal Research Center «Computer Science and Control» of the Russian Academy of Sciences, Moscow, Russia

<sup>III</sup> Moscow Institute of Physics and Technology, Moscow, Russia

**Abstract:** In this work we deal with the problem of recognition of printed document, captured by scanned devices and mobile phones. Recognition of documents' images rotated by 180 degrees, by known approaches involves orientation detection of image, then rotation if necessary, and the actual document image recognition in the correct orientation. The proposed approach based on convolutional neural network that is invariant to the angle of rotation by 180 degrees, eliminates the steps of orientation detection and image rotation. This speeds up the recognition process on mobile platforms, which performance is currently concedes to server and desktop platforms. Recognition of two data sets was considered: scanned images of structured national documents and public SmartDoc dataset, which contains images captured by mobile phones. For this datasets the accuracy of document recognition was estimated. The accuracy of the orientation detection using the proposed method on the considered stands is 100%, which exceeds the accuracy of the orientation detections of the methods described in the works from the list of references.

**Keywords:** document image recognition; orientation detection; rotation-invariant; image processing; mobile platforms.

DOI 10.14357/20718632190408

## Introduction

With development and spread of mobile technologies, not only the process of digitizing the document, but also its recognition is performed on smartphones. One of the initial steps before recognition is page orientation detection and image rotation of the image, if necessary. We suggest a document recognition method that is invariant to 180 degree rotation and avoids the steps of top-down direction orientation and image rotation, which is especially important for character recognition on mobile platforms, as they are limited in performance.

The problem of page orientation detection has long been known and there are various methods to solve it. We will consider orientation detection of a page, rotated 180 degrees (top down direction), provided that the text lines are slightly skewed in the case of a properly defined orientation.

A well-known approach to page orientation detection is to determine the ratio of ascenders and descenders (letter elements that go beyond the mean line and baseline - top and bottom of the letters 'w', 'e', 'r', 'u', 'o', 'a', 's', 'z', 'x', 'c', 'v', 'n', 'm') and compare with the known ratio for any language [1-4]. Other approaches of page orienta-

\* This work is partial financial support by Russian Foundation for Basic Research (projects 17-29-03170, 17-29-03236).

tion detection are recognition-driven [6] and neural network-driven [18, 19] methods.

In [1], a method was proposed, where ascenders and descenders are extracted using morphological operations. The error rate in the UW-I data set is 1 incorrectly defined orientation against 938 correctly defined ones. The orientation of the 41 documents cannot be determined due to the specifics of the method. Accuracy was 95.8%. Implementation of this method is available in Leptonica [8], the open source image-processing library. In [2], horizontal projections of the image are constructed in a special way, further analysis of which allows using the fact of difference between the number of ascenders and descenders. For all 226 images tested, the orientation was correctly determined.

Modifications of method based on ascenders and descenders are also applied in [3] and [4]. In [4] the method reaches an accuracy of 99.1% for dataset UW-I. It is resistant to various scanning resolutions and can reliably detect the page orientation of documents, displayed with a resolution of 150, 200, 300 and 400 dpi. Accuracy of 144 images was 100%.

The method, based on ascenders and descenders, is not applicable, when they are absent in the image, for example, in the case of some languages [5] or if a text consists only of capital letters. This method will also not be effective enough in the case of noisy or distorted data, which is common for images, obtained with a camera of mobile phone. The recognition-driven method was proposed in [6]; it should bypass above problems. For it to work, no additional data is required, except for those needed for the recognition process. In other words, orientation detection is embedded in the recognition process.

In the work [5], a method for the orientation detection and categorization of a language from four known (Arabic, Chinese, Korean, Roman) is presented. Since Chinese does not contain ascenders and descenders, like Roman languages, the orientation detection cannot be based on this information. The method suggested encodes orientation and language information into a vector, using the Vertical Component Run (VCR) density and distribution, which corresponds to the number of transitions from white to black in the vertical symbol section. Experiments on 492 test images of documents show that the average indicators for orienta-

tion detection and language categorization reach 97.56% and 99.59%, respectively.

In [6], the recognition-driven page orientation detection method (RD-POD) is described, which consists of several stages: binarization, selection of several lines in four possible orientation and character recognition of line.

Then the recognition quality in the four orientations is used as an assessment of whether the resulting orientations correspond to a language model, based on trigrams or vocabulary. The output of RD-POD is displaying the main orientation and pages, containing text in several orientations. The results are presented in the Tab. 1.

In [18] CNN was offered for determining the language and document orientation, based on a text string, using a sliding window. Before using CNN, text lines are extracted from the image, and, using a sliding window, are cut into small image patches. CNN is used to classify each patch by language and orientation. After that, the voting process is applied to obtain the results.

Offset neural network (ONN) for orientation identification in multinational languages was offered in [19]. It considers symbols that are symmetrical to rotation by 180 degrees, for example, "u" and "n", to reduce the number of errors. Such symbols are the problem for systems invariant to 180 degree rotation.

The page orientation techniques described above that use binary representations of documents are not applicable to color and grayscale images digitized by mobile cameras. This follows from the lack of quality of binarization due to specific distortions (glare, shape distortion of objects or defocusing [20]). Examples are shown on Fig. 1.

Tab. 1. The results of orientation detection, taken from [6]

Method	Dataset	Pages	Error	Rejection
RD-POD	Alice	44	0.000%	0.000%
	ICDAR07	40	0.000%	0.000%
	MARG	1553	0.064%	0.064%
	G1000	740	0.000%	0.676%
	UW3	1600	0.000%	0.313%
Bloomberg, 1995 [1]	UW1	979	0.110%	4.227%
Beusekom, 2009 [4]	UW1	979	0.928%	0.000%

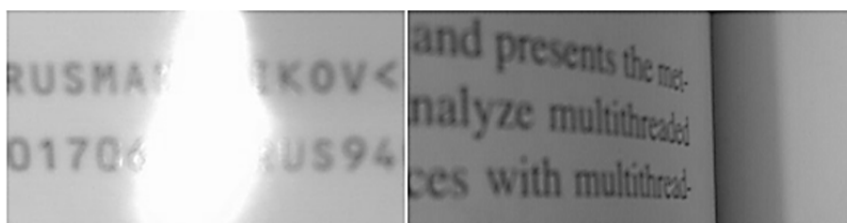


Fig. 1. Example of image distortions: glare (left); non-linear distortions (right)

In this work, we consider the problem of recognition of printed document, which image was captured with scanned devices and by camera of smartphone.

We suggest the approach to the document recognition, which eliminates the necessity of step of top-down orientation detection and rotation if needed so the processing time of the document is reduced.

For any computer platform, a significant part of the recognition process takes the actual character recognition. Currently, the conventional method of optical character recognition is artificial neural networks (ANN), which achieve the best results on public datasets for object classification [11-15].

Our method does not require orientation detection, since for character recognition we propose convolutional neural network (CNN) that are invariant to a 180 degree rotation. Recognition results are analyzed using frequency characteristics of language to obtain final result and detect orientation of document.

In this work, CNN is described, which is invariant to a 180 degree rotation, as well as several CNN applications for processing digitized documents.

## 1. Algorithm

### 1.1. Convolutional Neural Network for Character Recognition

For text recognition on document images, it is proposed to use a convolutional neural network, invariant to rotation by 180 degrees. The architecture of this CNN is presented in Tab. 2. The input for this network is the image of a separate character in grayscale, size - 15x19. To make this network invariant to rotation by 180 degrees, we used real-time augmentation. In general, it is a method of increasing variability of training dataset, which transforms it with different sets of distortions with random parameters on each learning iteration. In our case, we applied 180 degrees rotation to 50% of training set (selected randomly).

### 1.2. Extracting Attributes from a document with a Flexible Structure

Let's consider one of the applications, based on ANN and intended to extract attributes from a document with a flexible structure.

Tab. 2. Architecture of recognition CNN

№	Type	Parameters	Activation function
1	Conv	4 filters 3x3, stride 1x1, no padding	ReLU
2	Conv	8 filters 3x3, stride 1x1, padding 1x1	ReLU
3	Conv	12 filters 3x3, stride 2x2, padding 1x1	ReLU
4	Conv	12 filters 3x3, stride 1x1, padding 1x1	ReLU
5	Conv	16 filters 3x3, stride 1x1, padding 1x1	ReLU
6	Conv	24 filters 3x3, stride 2x2, padding 1x1	Tanh
7	Conv	16 filters 3x3, stride 1x1, padding 1x1	Tanh
8	Conv	12 filters 3x3, stride 1x1, padding 1x1	Tanh
9	Conv	8 filters 3x3, stride 1x1, padding 1x1	Tanh
10	Fully connected with softmax	55 neurons	

The document model with a flexible structure  $M$  is described as follows: there is a set of keywords  $W$  and fields for extracting attributes that are grouped into a set ordered according to the height of the strings:

$$S_1 = \{p_{11}, p_{12}, \dots, p_{1,k(1)}\},$$

$$S_2 = \{p_{21}, p_{22}, \dots, p_{2,k(2)}\},$$

$$\dots$$

$$S_i = \{p_{i1}, p_{i2}, \dots, p_{i,k(i)}\}$$

Each of the elements  $p_{ij} = \{xp^1_{ij}, yp^1_{ij}, xp^2_{ij}, yp^2_{ij}, tp_{ij}\}$  can be a word or a field. The coordinates (rectangle  $xp^1_{ij}, yp^1_{ij}, xp^2_{ij}, yp^2_{ij}$ ) of each element are known in advance. And the string value  $tp$  of each keyword is known.

For a pair of elements  $p = \{xp^1, yp^1, xp^2, yp^2, tp\}$  and  $q = \{xq^1, yq^1, xq^2, yq^2, tq\}$  define relations:

$$\text{left}(p, q) \text{ if } xp^2 > xq^1,$$

$$\text{above}(p, q) \text{ if } yp^2 < yq^1$$

For a pair of strings  $S_i = \{p_{i1}, p_{i2}, \dots\}$  and  $S_r = \{p_{r1}, p_{r2}, \dots\}$  define relation:

$$\text{above}(S_b, S_r) \text{ if } (\forall p_i \in S_b, \forall p_r \in S_r) \text{ above}(p_b, p_r)$$

The following relations should be performed in the model:

$$(\forall t, r : t < r) \text{ above}(S_b, S_r),$$

$$(\forall p_i \in S, \forall p_r \in S_r : t < r) \text{ left}(p_i, p_r) \quad (1)$$

Let us consider an image, part of which is occupied by the document, and set of recognized words  $W_1$ , extracted from this image. Assuming that the document is a known model  $M$ , we find a subset of words  $W_1 \subset W$  that best matches the model, i.e., we find the maximum number of words  $w = \{xw^1, yw^1, xw^2, yw^2, tw\} \in W_1$ , each of which is associated with an element of the model  $p = \{xp^1, yp^1, xp^2, yp^2, tp\} \in M$ , so that the words  $tw$  and  $tp$  are close in measure  $d$ , for example, the Levenshtein metric. Moreover, for each pair of words  $w_1, w_2 \in W_1$ , relations (1) are fulfilled. Placing the recognized words requires a search, the peculiarity of which is the elimination of the conflict outliers of similar words, associated primarily with recognition errors. After establishing a match between  $W_1$  and  $M$  (model bindings), field values may be extracted, using boundaries of the associated keywords.

The sizes of the elements and the distances between the elements are not constant, for real documents the relations between the corresponding distances can exceed 100%.

The proposed algorithm (A) for extracting attributes from a scanned image of a document with a flexible structure is as follows:

- 1) image normalization (transform to the single-channel image, deskew image),
- 2) finding word boundaries using morphological operations of erosion and dilatation,
- 3) segmentation of the found words into characters, using an ANN that is invariant to a rotation by 180 degrees,
- 4) recognition of each found character, using the ANN that is invariant to rotation by 180 degrees,
- 5) control the presence of unique keywords from set  $W$
- 6) clustering words into strings by the nearest neighbors algorithm,
- 7) binding and extracting fields as written above.
- 8) getting results as extracted attributes and document orientation

Formation of word recognition results is shown in Fig. 2.

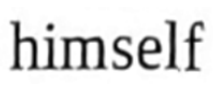
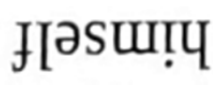
Comparison of the recognized word  $w = \{xw^1_{ij}, yw^1_{ij}, xw^2_{ij}, yw^2_{ij}, tw\}$  and words from the model  $p_{ij} = \{xp^1_{ij}, yp^1_{ij}, xp^2_{ij}, yp^2_{ij}, tp\}$  must be performed twice, using two metrics. First is the Levenshtein metric  $d(tw, tp)$ , using which character sequences  $tw_1, tw_2, \dots, tw_n$ , and  $tp_1, tp_2, \dots, tp_m$  are compared. Second metric is  $d_{rev}(tw, tp) = d(tp, tw)$ , using which sequences  $tw_1, tw_2, \dots, tw_n$  and  $tp_m, tp_{m-1}, \dots, tp_1$  are compared.

To control the presence of unique keywords, it is possible to use distance

$$d_{inv}(tw, tp) = d_{rev}(tw, tp) + d(tw, tp).$$

Obviously,  $d_{inv}$  is a metric that is invariant to the top-down orientation of the original word images when comparing recognized words.

Fig.2. Examples of word recognition using the ANN, invariant to rotate by 180 degrees

Image	Result of recognition
	HIMSELF
	FLESMIH

To take into account the fact of rotation set by the  $d_{rev}$  metric, it is necessary to transform the coordinates of each recognized word  $w = \{xw^1_{ij}, yw^1_{ij}, xw^2_{ij}, yw^2_{ij}, tw\}$ :

$$w_{rot} = \{Wid-xw^2_{ij}, Hei-yw^2_{ij}, Wid-xw^1_{ij}, Hei-yw^1_{ij}, tw\}, \quad (2)$$

where  $Wid$  and  $Hei$  – sizes of the recognized image.

Thus, when using ANN, which is invariant to 180 degree rotation, and  $d_{inv}$  and  $d_{rev}$  metrics, the keyword binding, regarding relations (1), is carried out without the need to detect orientation of the set of recognized.

### 1.3. Recognition a Set of words from a document with an Arbitrary Structure

Another application is designed to recognize a set of words from a document with an arbitrary structure.

The algorithm (B) for recognition of a set of words from a photo of document with an arbitrary structure, based on the use of ANN, invariant to a 180 degree rotation is as follows:

- 1) selection of the document boundaries in the photo and projective rectification,
- 2) steps 2-4 from algorithm (A),
- 3) getting results as set of words and document orientation.

In the case of original image has been rotated by 180 degree, only a reversal of the character order in the words is required. In other words, each word  $tw_1, tw_2, \dots, tw_n$  must be represented as  $tw_n, \dots, tw_2, tw_1$ . To change the order of the words (and strings), the

word coordinate transformation (2) is required.

At the same time, to determine whether a reordering of words is required (whether the original image was rotated – top-down orientation detection), it is necessary to check the obtained recognition results to match the language model, using frequency characteristics. In addition, in the case of rotating the original image by 180°, such pairs of characters as “d” and “p” or “6” and “9” are considered as equal.

## 2. Results

In the experiments we determined the accuracy of document orientation detection and the recognition accuracy. Recognition accuracy was determined by the formula, presented in [17]:

$$\text{Character accuracy} = \frac{n - \#errors}{n},$$

where  $n$  - total number of characters in ground truth data,  $\#errors$  - number of errors; each insertion, deletion or replacement of a character is considered as an error.

### 2.1. Experiment of recognition of National Structured Document Dataset (PIT - Personal Income Tax Document)

For the experiment, a dataset from 2000 marked scans of documents with a resolution greater than 1000x1500 pixels was prepared: 1000 scans are in the correct orientation; the rest are their copies, rotated by 180 degrees. Example of PIT document with found fields is presented in Fig. 3.

**СПРАВКА О ДОХОДАХ ФИЗИЧЕСКОГО ЛИЦА**

Приложение № 1  
к приказу ФНС России  
от 30.10.2015 № ММВ-7-  
11/485@

за  год №  от

Признак  номер корректировки  в ИФНС (код)

**Форма 2-НДФЛ**  
**Код по КНД 1151078**

**1. Данные о налоговом агенте**

Код по ОКТМС  Телефон  ИНН  КПП

Налоговый агент

**2. Данные о физическом лице - получателе дохода**

ИНН в Российской Федерации  ИНН в стране гражданства

Фамилия  Имя  Отчество

Статус налогоплательщика  Дата рождения  Гражданство (код страны)

Код документа, удостоверяющего личность:  Серия и номер документа

Адрес места жительства в Российской Федерации: Почтовый индекс  Код субъекта

Район  Город  Населенный пункт

Улица  Стрит  Дом  Корпус  Квартира

Код страны проживания:  Адрес

**3. Доходы, облагаемые по ставке**

Fig. 1. Example of a PIT document and found fields

The recognition was reproduced in accordance with algorithm (A), described above. The results of experiments on Results of experiments (PIT) correctly oriented images and images, rotated by 180 degrees, are presented in the Tab. 3.

**2.2. SmartDoc Dataset Recognition Experiment**

For experiments with document recognition, the public SmartDoc dataset was chosen [16]: Smartphone document capture and OCR - images of documents, captured using a mobile device with different angles of shooting and distortion, which contains 8470 images. Since in the SmartDoc dataset images are in horizontal orientation, which we do not define, all the images are preoriented to the correct document orientation or to the orientation, rotated by 180 degrees; we get 16940 images.

Recognition was performed using the algorithm (B), invariant to the page orientation. The results of experiments of images in default orientation, and images rotated by 180 degrees, are presented in the Tab. 4.

**3. Conclusion**

In this article, we described an approach to the recognition of printed documents, which allows extracting data from documents with different orientations (default and rotated by 180 degrees) without rotating the recognized image. This reduces time of document’s image processing that

can be particularly useful for processing an image of document on mobile platforms.

Our approach is based on convolutional neural network (CNN) invariant to 180 degree rotation. The recognition accuracy of rotated by 180 degrees images differs from the recognition accuracy of images in default orientation by 1.25% on own dataset of national structured documents (Russian PIT), consisting of scanned images, and 2.73% on SmartDoc dataset, consisting of images, obtained using a smartphone camera.

According to results of experiments, with almost the same recognition accuracy of images in default orientation and 180 degree rotated images we have approach of document recognition that skips the step of orientation detection before recognition because of using CNN invariant to 180 degree rotation.

The accuracy of orientation detection on the used datasets is 100%, which corresponds to the accuracy of the known methods. Our approach can be simply and successfully used for hypothetical case of recognition of document page containing text in mixed top-down orientations, while other methods are difficult to adapt.

**References**

1. D. S. Bloomberg, G. E. Kopec, and L. Dasari, “Measuring document image skew and orientation,” in Proc. SPIE Document Recognition II, pp. 302–316, (San Jose, CA, USA), Feb. 1995.

Tab. 3. Results of experiments (PIT)

	Orientation detection accuracy	Recognition accuracy
Images in default (0°) orientation	100	96.88%
Images rotated by 180°	100	95.63%

Table 4/ Results of experiments (SmartDoc)

	Orientation detection accuracy	Recognition accuracy
Images in default (0°) orientation	100	79.86%
Images rotated by 180°	100	77.13%

2. R. S. Caprari, "Algorithm for text page up/down orientation determination," *Pattern Recognition Letters* 21(4), pp. 311–317, 2001
3. B. T. Avila and R. D. Lins, "A fast orientation and skew detection algorithm for monochromatic document 'images,'" in *DocEng '05: Proc. ACM Symposium on Document Engineering*, 2005, pp. 118–126. doi:10.1145/1096601.1096631
4. J. van Beusekom, F. Shafait, T. M. Breuel, "Resolution independent skew and orientation detection for document images", *Proc. SPIE 7247, Document Recognition and Retrieval XVI, 72470K* (19 January 2009); doi: 10.1117/12.807735
5. S. Lu, C. L. Tan, "Automatic document orientation detection and categorization through document vectorization". In: K. Nahrstedt, M. Turk, Y. Rui, W. Klas, K. Mayer-Patel (eds.) *Proc. 14th ACM International Conference on Multimedia* October 23-27, 2006, Santa Barbara, CA, USA. pp. 113-116.
6. Y. Rangoni, F. Shafait, J. van Beusekom & T. M Breuel, "Recognition driven page orientation detection". 2009 16th IEEE International Conference on Image Processing (ICIP). doi:10.1109/icip.2009.5413722
7. V. Konya, S. Eickeler & C. Seibert, "Fast seamless skew and orientation detection in document images". 20th International Conference on Pattern Recognition, 2010. doi:10.1109/icpr.2010.474
8. URL: <http://www.leptonica.com/>
9. E. Limonova, D. Ilin, D. Nikolaev, "Improving Neural Network Performance on SIMD Architectures". *Proc. SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015)*, 98750L (8 December 2015); doi:10.1117/12.2228594
10. V. Gayer, A. V. Sheshkus, Y. S. Chernyshova, "Effective real-time augmentation of training dataset for the neural networks learning" *ICMV-2018*
11. L. Wa, M. Zeiler, S. Zhang, Y. LeCun, and R. Fergus, "Regularization of Neural Networks using DropConnect" *Proc. ICML'13 30th International Conference on International Conference on Machine Learning*, vol. 28, 2013, pp. 1058-1066.
12. B. Graham, "Fractional max-pooling." arXiv preprint arXiv:1412.6071 (2014).
13. DA. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," arXiv:1511.07289, 2015.
14. J. Zhao, M. Mathieu, R. Goroshin, and Y. LeCun, "Stacked What-Where Auto-encoders," arXiv:1506.02351.
15. CY. Lee, P.W. Gallagher, and Z. Tu, "Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree," arXiv:1509.08985.
16. J.-C. Burie, J. Chazalon, M. Coustaty, S. Eskenazi, M. M. Luqman, M. Mehri, N. Nayef, J.-M. OGIER, S. Prum and M. Rusinol, "ICDAR2015 Competition on Smartphone Document Capture and OCR (SmartDoc)", In 13th International Conference on Document Analysis and Recognition (ICDAR), 2015.
17. L. Blando, J. Kanai, and T. Nartker, "Prediction of OCR accuracy using simple image features," in *International Conference on Document Analysis and Recognition*, vol. 1, 1995, pp. 319–322.
18. Chen, L., Wang, S., Fan, W., Sun, J., & Satoshi, N, "Deep learning based language and orientation recognition in document analysis". 13th International Conference on Document Analysis and Recognition (ICDAR), 2015. doi:10.1109/icdar.2015.7333799
19. R. Wang, S. Wang, & J. Sun, "Offset Neural Network for Document Orientation Identification". 13th IAPR International Workshop on Document Analysis Systems (DAS). 2018. doi:10.1109/das.2018.12
20. K. Bulatov, V. Arlazarov, T. Chernov, O. Slavin, and D. Nikolaev, "Smart IDReader: Document recognition in video stream" *The 14th IAPR International Conference on Document Analysis and Recognition (ICDAR 2017), Workshops and Tutorials: November 9-12, Kyoto, Japan, 2017* – p. 39-44. ISSN: 2379-2140 <http://ieeexplore.ieee.org/document/8270294/>

**Andreeva E. I.** Smart Engines Service LLC, Prosp. 60-letiya Oktyabrya, 9, Moscow, 117312, Russia e-mail: andreeva@phystech.edu

**Arlazarov V. V.** Candidate of Science (PhD) in technology, Computing Center, Federal Research Center «Informatics and Management» of the Russian Academy of Sciences, Prosp. 60-letiya Oktyabrya, 9, Moscow, 117312, Russia, e-mail: vva@smartengines.biz

**Gayer A. V.** Smart Engines Service LLC, Prosp. 60-letiya Oktyabrya, 9, Moscow, 117312, Russia, e-mail: gayer.alexandr@yandex.ru

**Dorokhov E. P.** Smart Engines Service LLC, Prosp. 60-letiya Oktyabrya, 9, Moscow, 117312, Russia, e-mail: eudorokhov@gmail.com

**Sheshkus A. V.** Federal Research Center «Informatics and Management» of the Russian Academy of Sciences, Prosp. 60-letiya Oktyabrya, 9, Moscow, 117312, Russia, e-mail: astdcall@gmail.com.

**Slavin O. A.** Doctor of Science in technology, Computing Center, Federal Research Center «Informatics and Management» of the Russian Academy of Sciences, Prosp. 60-letiya Oktyabrya, 9, Moscow, 117312, Russia, e-mail: OSlavin@isa.ru (corresponding author).