

Algorithm of Text Recognizing in Ukrainian on the Video Mode

Kateryna Tymoshenko, Victoria Vysotska

Lviv Polytechnic National University, Lviv, Ukraine
Victoria.A.Vysotska@lpnu.ua

Abstract. OCR technology for recognizing text in Ukrainian is imperfect. The Android operating system contains a character repository of many languages, but not the Ukrainian. Therefore, we propose a method and algorithm for recognizing text in Ukrainian and modifying it for video mode.[1].

1 Introduction

To solve this problem, we create:

- XML file, which contains uppercase and lowercase Ukrainian characters;
- a class that recognizes text and places the recognized character in a temporary buffer;
- a class that modifies the work of the previous class with an array of frames and recognizes text.

2 Text Recognition Algorithm in Ukrainian

For recognizing Ukrainian text, first of all the XML file with uppercase and lowercase Ukrainian characters has been created. Each character from XML file is assigned its own name [2]. Example of XML file:

```
<resources>
  <string name="A">A</string>
    <string name="a">a</string>
  <string name="B">B</string>
    <string name="б">б</string>
  <string name="B">B</string>
    <string name="в">в</string>
</resources>
```

The next step is to connect the class to recognize the text in the image. A character is taken from the screen and compared to a character from an XML file (Fig. 1).

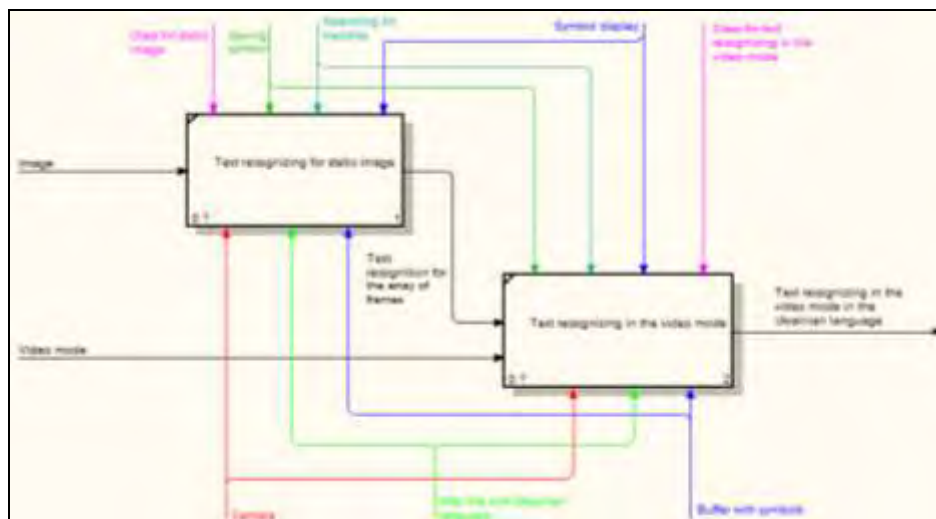


Fig. 1. Flowchart for identifying text in Ukrainian

To compare characters from an image and from an XML file, the character in the image is converted to a specific standard. Namely, the font and its size changes. In our case, it's a Courier New 10pt (Fig. 2).

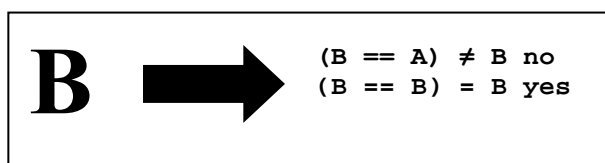


Fig. 2. Conversion a character symbol to the standard

Then the symbol from the image and the file are compared. For this purpose, a script is used to compare characters from OCR technology [1]. Once a character is found, it is stored in a temporary buffer that sequentially accumulates characters.

3 The Result of theA

After the recognition of all the text is completed, the result is displayed on the screen as a frame, inside which is the recognized text (Fig. 3).

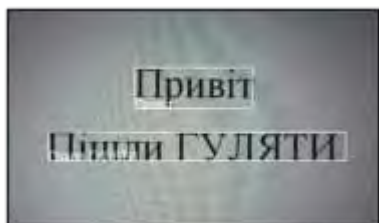


Fig. 3. Text recognition

In the future, for the text to be recognized when the images are changed, the temporary buffer is released, and the new image is processed in a new way. To do this, the main class will connect to a class that works with an array of frames. The result is recognition of Ukrainian text when changing frames in video mode.

4 Testing of developed algorithm

Let's compare the quality and speed of the developed algorithm with all known OCRs. The research of the work of these algorithms was carried out under the same conditions. Speed of algorithms were tested on depends of changing of illuminations, counts of characters and both conditions. For each research were 20 tests.

As graph below shows the OCR algorithm requires more time and frames for clear text recognition. In average the OCR algorithm need 1 frame more than developed algorithm. (Fig. 4)

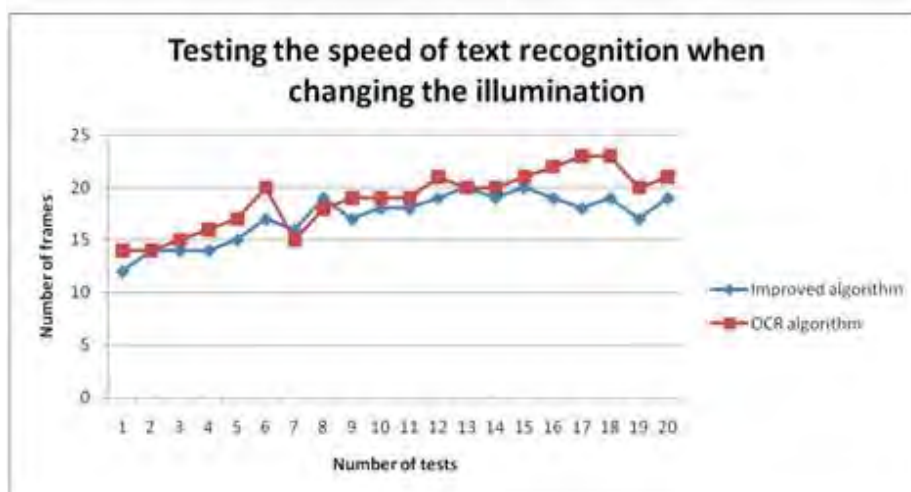


Fig. 4. Testing text recognizing when changing the illumination

The next graph shows how much frames both of algorithms need for recognizing text with different count of characters. The result is the OCR algorithm requires more time and frames to recognize the text, the deviation occurs at 1.5 - 2 frames. (Fig.5)

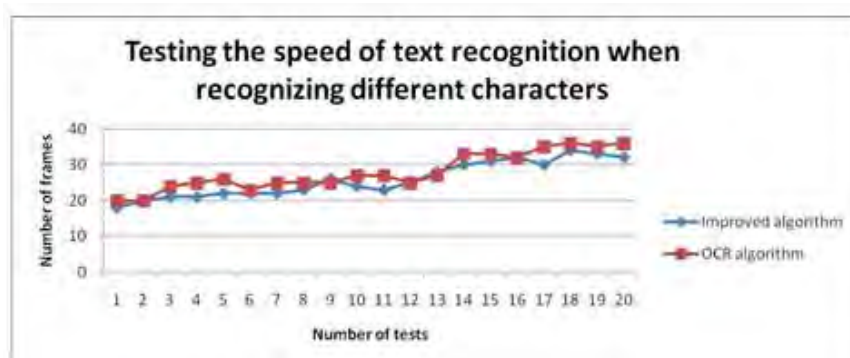


Fig. 5. Testing text recognizing when changing the number of symbols

And the last one shows how speed of algorithms depend of both of conditions – illumination and count of characters. Due to this graph we can conclude that the developed algorithm yields an increase in productivity by 4-5%, and if more precisely, the recognition of the text goes to 1-2 frames less than the OCR algorithm.

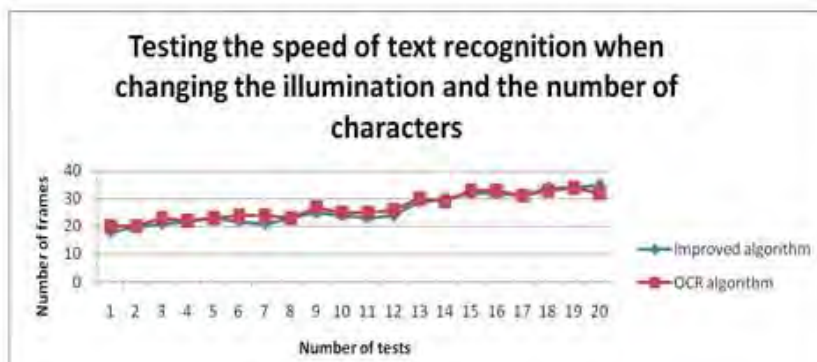


Fig. 6. Testing text recognizing when changing the number of symbols and changing the illumination

5 Conclusions

The results of algorithm and software development are developing an XML file that contains uppercase and lowercase characters; develop a class that recognizes text and places a recognized character in a temporary buffer, after processing the characters,

the text is displayed; develop a class that uses the first class as the basis for recognizing text when frames change.

The analysis showed that there is a need to develop high-speed algorithms for the recognition of Ukrainian texts in order to implement them in mobile devices.

The program is written in the Java language. Consists of turning on the camera and prompts for the user.

References

1. De Campos T. E. "Character Recognition in Natural Images in the Proceedings of the International Conference on Computer Vision Theory and Applications" – Lisbon, Portugal, February – 2009.
2. Yokobayashi M. "Binarization and recognition of degraded characters using a maximum separation axis in the color space" / Wakahara T. – Vol. 2nd – 2006 – pp. 885–888.
3. Pan Y. "Text localization in natural scene images based on a conditional random field in the International Conference on Document Analysis and Recognition" / Hou X., Liu C. - 2009.
4. Demchuk, A., Lozynska, O.: The Typhlocomments Rules for Audiodescription System of the Video Content Formation for People with Visual Impairments. In: Computational Linguistics and Intelligent Systems, COLINS, 2, 53-59. (2018)
5. Lytvyn, V., Vysotska, V., Mykhailyshyn, V., Rzheuskyi, A., Semianchuk, S.: System Development for Video Stream Data Analyzing. In: Advances in Intelligent Systems and Computing, 1020, 315-331. (2020)
6. Литвин В. В., Висоцька В. А., Михайлишин В. Ю., Сем'янчук С. О. Розроблення інформаційної системи аналізу даних відеопотоку // Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту : збірка наукових праць Міжнародної наукової конференції (с. Залізний Порт, 21–25 травня 2019 р.). – 2019. – С. 94–97
7. Veres, O., Rishnyak, I., Rishniak, H.: Application of Methods of Machine Learning for the Recognition of Mathematical Expressions. In: Computational linguistics and intelligent systems, COLINS, 378-389. (2019)
8. Bakumenko, N., Strilets, V., Ugryumov, M.: Application of the C-Means Fuzzy Clustering Method for the Patient's State Recognition Problems in the Medical Monitoring System. In: Computational linguistics and intelligent systems, COLINS, 218-227. (2019)
9. Dovbysh, A., Shelehov, I., Pylypenko, S., Berest, O.: Estimation of Informativeness of Recognition Signs at Extreme Information Machine Learning of Knowledge Control System. In: Computational linguistics and intelligent systems, COLINS, 143-152. (2019)
10. Dovbysh, A., Aliexsieiev, V.: Embedding Speech Recognition Tools for Custom Software: Engines Overview. In: Computational Linguistics and Intelligent Systems, COLINS, 2, 114-121. (2018)
11. Lytvyn, V., Vysotska, V., Pukach, P., Bobyk, I., Uhryn, D.: Development of a method for the recognition of author's style in the Ukrainian language texts based on linguometry, stylemetry and glotochronology. In: Eastern-European Journal of Enterprise Technologies, 4(2-88), 10-19. (2017)
12. Lytvyn, V., Peleshchak, I., Vysotska, V., Peleshchak, R.: Satellite spectral information recognition based on the synthesis of modified dynamic neural networks and holographic

- data processing techniques. In: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 330-334. (2018)
13. Zdebskyi, P., Vysotska, V., Peleshchak, R., Peleshchak, I., Demchuk, A., Krylyshyn, M.: An Application Development for Recognizing of View in Order to Control the Mouse Pointer. In: CEUR Workshop Proceedings, Vol-2386, 55-74. (2019)
 14. Shu, C., Dosyn, D., Lytvyn, V., Vysotska V., Sachenko, A., Jun, S.: Building of the Predicate Recognition System for the NLP Ontology Learning Module. In: International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS, 2, 802-808. (2019)
 15. Kazarian, A., Teslyuk, V., Tsmots, I., Tykhan, M.: Implementation of the Face Recognition Module for the “Smart” Home Using Remote Server. In: Advances in Intelligent Systems and Computing III, AISC 871, Springer, 17–27. (2019)
 16. Komar, M., Golovko, V., Sachenko, A., Bezobrazov, S.: Development of neural network immune detectors for computer attacks recognition and classification. In: International Conference on Intelligent Data Acquisition and Advanced Computing Systems, IDAACS, 665-668. (2013)
 17. Paliy, I., Turchenko, V., Koval, V., Sachenko, A., Markowsky, G.: Approach to recognition of license plate numbers using neural networks. In: International Conference on Neural Networks – Conference Proceedings, 2965-2970. (2004)
 18. Furgala, Yu.M., Rusyn, B.P.: Peculiarities of melin transform application to symbol recognition. In: 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET, 251-254. (2018)
 19. Kapustiy, B.E., Rusyn, B.P., Tayanov, V.A.: Peculiarities of application of statistical detection criteria for problems of pattern recognition. In: Journal of Automation and Information Science, 37(2), 30-36. (2005)
 20. Rusyn, B., Kosarevych, R., Lutsyk, O., Korniy, V.: Segmentation of atmospheric clouds images obtained by remote sensing. In: 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET, 213-216. (2018)
 21. Kapustiy, B.O., Rusyn, B.P., Tayanov, V.A.: A new approach to determination of correct recognition probability of set objects. In: *Upravlyaushchie Sistemy i Mashyny*, 2, 8-12. (2005)
 22. Rusyn, B., Prudyus, I., Ostap, V.: Fingerprint image enhancement algorithm. In: The Experience of Designing and Application of CAD Systems in Microelectronics, Proceedings of the 6th International Conference ,CADSM, 193-194. (2001)
 23. Rusyn, B., Torska, R., Kobasyar, M.: Application of the cellular automata for obtaining pitting images during simulation process of their growth. In: Advances in Intelligent Systems and Computing, 242, 299-306. (2014)
 24. Varetskyy, Y., Rusyn, B., Molga, A., Ignatovych, A.: A new method of fingerprint key protection of grid credential. In: Advances in Intelligent and Soft Computing, 84, 99-103. (2010)
 25. Vysotska V. Linguistic analysis and modelling semantics of textual content for digest formation / Victoria Vysotska, Lyubomyr Chyrun // MEST Journal (Management Education Science & Society Technologie). – Vol.3 No.1. – PP. 127-148 [Online]. – ISSN 2334-7171, ISSN 2334-7058 (Online), DOI 10.12709/issn.2334-7058. This issue: DOI 10.12709/mest.02.02.02.0. – Режим доступу: http://mest.meste.org/MEST_1_2015/Sadrzaj_eng.html
http://mest.meste.org/MEST_1_2015/5_15.pdf.

26. Berko A. Linguistic Analysis for the Textual Commercial Content / Andriy Berko, Victoria Vysotska, Lyubomyr Chyrun // Computer Science and Information Technologies: Proc. of the IX-th Int. Conf. CSIT'2014, 18-22 November, 2014, Lviv, Ukraine..– Lviv: Publishing Lviv Polytechnic, 2014.– P.11-14.
27. Shestakevych T. Modelling of semantics of natural language sentences using generative grammars / Tetiana Shestakevych, Victoria Vysotska, Lyubomyr Chyrun, Liliya Chyrun // Computer Science and Information Technologies: Proc. of the IX-th Int. Conf. CSIT'2014, 18-22 November, 2014, Lviv, Ukraine..– Lviv: Publishing Lviv Polytechnic, 2014.– P.19-22.
28. Andrunyk V. The Peculiarities of Electronic Digest Formation / Vasyl Andrunyk, Victoria Vysotska, Lyubomyr Chyrun // Computer Science and Information Technologies: Proc. of the IX-th Int. Conf. CSIT'2014, 18-22 November, 2014, Lviv, Ukraine..– Lviv: Publishing Lviv Polytechnic, 2014.– P.25-28.
29. Kis I. Features of the Content-Analysis Method in Processing Online Newspaper Articles / Iaroslav Kis, Victoria Vysotska, Liliya Chyrun, Vasyl Foltovych // Computer Science and Information Technologies: Proc. of the IX-th Int. Conf. CSIT'2014, 18-22 November, 2014, Lviv, Ukraine..– Lviv: Publishing Lviv Polytechnic, 2014.– P.39-42.
30. Vysotska V. Generative regular grammars application to modeling the semantics of sentences in natural language / Victoria Vysotska // Комп'ютерні системи проектування. Теорія і практика, Вісник Національного університету "Львівська політехніка" № 808,- Львів 2014 – Стр.43-56.
31. Vysotska V. Linguistic Analysis of Textual Commercial Content for Information Resources Processing / Victoria Vysotska // Proceedings of the XIIIth International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET'2016). - February 23–26, 2016. Lviv–Slavske, Ukraine.- P. 709-713.
32. Vysotska V. Analysis of business processes in electronic content-commerce systems / V. Vysotska, L. Chyrun, P. Kozlov // ECONTECHMOD. An international quarterly journal on economics in technology, new technologies and modeling processes. - Polish Academy of Sciences, University of Engineering and Economics in Rzeszyw, Lviv Polytechnic National University, University of Life Sciences in Lublin, Faculty of Production Engineering. - Vol. 5, No 1. - LUBLIN–RZESZYW,2016. – pp. 111-125. - ISSN 2084-5715. - <http://www.pan-ol.lublin.pl/wydawnictwa/Econtechmod.html>.
33. Chyrun Lyubomyr. Specifics Informational Resources Processing for Textual Content Linguistic Analysis / Lyubomyr Chyrun, Victoria Vysotska, Vasyl Lytvyn // Proceeding of XIIth International Conference of Perspective Technologies and Methods in MEMS Design, MEMSTECH 2016. – 20-24 April, 2016, Lvi-Polyana, Ukraine. – Lviv Politechnic Publishing House. – P. 214-219.
34. Chyrun L. Informational resources processing intellectual systems with textual commercial content linguistic analysis usage constructional means and tools development / L. Chyrun, V. Vysotska, I. Kozak // Econtechmod : an international quarterly journal on economics in technology, new technologies and modelling processes. – Lublin ; Rzeszow, 2016. – Volum 5, number 2. – P. 85–94. – Bibliography: 60 titles.
35. Lytvyn Vasyl. Content Linguistic Analysis Methods for Textual Documents Classification / Vasyl Lytvyn, Victoria Vysotska, Oleh Veres, Ihor Rishnyak, Halya Rishnyak // Computer Science and Information Technologies: Proc. of the XI-th Int. Conf. CSIT'2016, 6-10 September, 2016, Lviv, Ukraine..– Lviv: Lviv Polytechnic Publishing House, 2016.– P.190-192.

36. Lytvyn V. Intelligent System Structure for Web Resources Processing and Analysis / V. Lytvyn, V. Vysotska, L. Chyrun, A. Smolarz, O. Naum // 1st International Conference Computational Linguistics and Intelligent Systems, COLINS'2017. – 21 April 2017, Kharkiv. – P. 56-74.
37. Lytvyn V. A Method of Construction of Automated Basic Ontology / V. Lytvyn, V. Vysotska, W. Wojcik, D. Dosyn // 1st International Conference Computational Linguistics and Intelligent Systems, COLINS'2017. – 21 April 2017, Kharkiv. – P. 75-83.
38. Lytvyn V. Development of a method for the recognition of author's style in the ukrainian language texts based on linguometry, stylemetry and glottochronology / V. Lytvyn, V. Vysotska, P. Pukach, I. Bobyk, D. Uhryn // Eastern-European Journal of Enterprise Technologies. – №4/2(88). – Харків 2017. – P. 10-18. – ISSN 1729-3774.
39. Korobchinsky Maksym. Peculiarities of Content Forming and Analysis in Internet Newspaper Covering Music News / Maksym Korobchinsky, Victoria Vysotska, Liliya Chyrun, Lyubomyr Chyrun // Computer Science and Information Technologies: Proc. of the XII-th Int. Conf. CSIT'2017. – 05-08 September, 2017, Lviv. – P. 52-57.
40. Naum O. Intellectual System Design for Content Formation / O. Naum, L. Chyrun, O. Kanishcheva, V. Vysotska // Computer Science and Information Technologies: Proc. of the XII-th Int. Conf. CSIT'2017. – 05-08 September, 2017, Lviv. – P. 131-138.
41. Lytvyn V. Application of Sentence Parsing for Determining Keywords In Ukrainian Texts / V. Lytvyn, V. Vysotska, D. Dosyn, R. Holoschuk, Z. Rybchak // Computer Science and Information Technologies: Proc. of the XII-th Int. Conf. CSIT'2017. – 05-08 September, 2017, Lviv. – P. 326-331.
42. Lytvyn Vasyly. The Contextual Search Method Based on Domain Thesaurus / Vasyly Lytvyn, Victoria Vysotska, Yevhen Burov, Oleh Veres, Ihor Rishnyak // Advances in Intelligent Systems and Computing. – Vol. 689. – ISSN 2194-5357 ISSN 2194-5365 (electronic). - ISBN 978-3-319-45990-5 ISBN 978-3-319-45991-2 (eBook). –DOI 10.1007/978-3-319-45991-2. - Springer International Publishing AG 2017. –PP. 310-319.
43. Kanishcheva Olga. Method of Integration and Content Management of the Information Resources Network / Olga Kanishcheva, Victoria Vysotska, Lyubomyr Chyrun, Aleksandr Gozhij // Advances in Intelligent Systems and Computing. – Vol. 689. – ISSN 2194-5357 ISSN 2194-5365 (electronic). - ISBN 978-3-319-45990-5 ISBN 978-3-319-45991-2 (eBook). –DOI 10.1007/978-3-319-45991-2. - Springer International Publishing AG 2017. –PP. 204-216.
44. Jun Su. Information resources processing using linguistic analysis of textual content / Jun Su, Victoria Vysotska, Anatoliy Sachenko, Vasyly Lytvyn, Yevhen Burov // Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2017 9th IEEE International Conference. – 21-23 Sept. 2017, Bucharest, Romania. – P. 573 – 578.
45. Lytvyn V. The Risk Management Modelling in Multi Project Environment / V. Lytvyn, V. Vysotska, O. Veres, I. Rishnyak, H. Rishnyak // Computer Science and Information Technologies: Proc. of the XII-th Int. Conf. CSIT'2017. – 05-08 September, 2017, Lviv. – P. 32-35.
46. Lytvyn, V., Peleshchak, I., Peleshchak, R.: Increase the speed of detection and recognition of computer attacks in combined diagonalized neural networks. In: 4th International Scientific-Practical Conference Problems of Infocommunications Science and Technology, PIC S and T, 152-155. (2018)

47. Zhezhnych, P., Markiv, O.: Recognition of tourism documentation fragments from web-page posts. In: 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET, 948-951. (2018)
48. Boyko, N., Basystiuk, O., Shakhovska, N.: Performance Evaluation and Comparison of Software for Face Recognition, Based on Dlib and Opencv Library. In: International Conference on Data Stream Mining and Processing, DSMP, 478-482. (2018)
49. Martsyshyn, R., Medykovskyy, M., Sikora, L., (...), Lysa, N., Yakymchuk, B.: Technology of speaker recognition of multimodal interfaces automated systems under stress. In: International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM, 447-448. (2013)
50. Shanidze, O., Petrasova, S.: Extraction of semantic relations from Wikipedia text corpus. In: Computational linguistics and intelligent systems, COLINS, 2, 74-75. (2019)
51. Razno, M.: Machine learning text classification model with NLP approach. In: Computational linguistics and intelligent systems, COLINS, 2, 71-73. (2019)
52. Manuilov, I., Petrasova, S.: Method for paraphrase extraction from the news text corpus. In: Computational linguistics and intelligent systems, COLINS, 2, 69-70. (2019)
53. Lytvynenko, J.: Identify of the substantive, attribute, and verb collocations in russian text. In: Computational linguistics and intelligent systems, COLINS, 2, 66-68. (2019)
54. Liutenko, I., Kurasov, O.: Consideration of the software tests quality evaluation problem. In: Computational linguistics and intelligent systems, COLINS, 2, 62-65. (2019)
55. Drobot, T.: Data-to-text generation for domain-specific purposes. In: Computational linguistics and intelligent systems, COLINS, 2, 60-61. (2019)
56. Chuiko, V., Khairova, N.: Semantic similarity identification for short text fragments. In: Computational linguistics and intelligent systems, COLINS, 2, 57-59. (2019)
57. Bobrovnyk, K.: Automated building and analysis of Ukrainian Twitter corpus for toxic text detection. In: Computational linguistics and intelligent systems, COLINS, 2, 55-56. (2019)
58. Bilova, M., Trehubenko, O.: Study of software systems usability used for customers loyalty identification. In: Computational linguistics and intelligent systems, COLINS, 2, 46-54. (2019)
59. Puzik, O.: Intelligence knowledge-based system based on multilingual dictionaries. In: Computational linguistics and intelligent systems, COLINS, 2, 39-44. (2019)
60. Berko, A.: Knowledge-based Big Data cleanup method. In: Computational linguistics and intelligent systems, COLINS, 2, 14-21. (2019)
61. Klyushin, D., Lyashko, S., Zub, S.: A (n) Assumption in machine learning. In: Computational linguistics and intelligent systems, COLINS, 2, 32-38. (2019)
62. Grabar, N., Hamon, T.: WikiWars-UA: Ukrainian corpus annotated with temporal expressions. In: Computational linguistics and intelligent systems, COLINS, 2, 22-31. (2019)
63. Kirichenko, L., Radivilova, T., Tkachenko A.: Comparative Analysis of Noisy Time Series Clusterin. In: Computational linguistics and intelligent systems, COLINS, 184-196. (2019)
64. Kazarian, A., Kunanets, N., Pasichnyk, V., Veretennikova, N., Rzheuskyi, A.: Complex Information E-Science System Architecture based on Cloud Computing Model. In: Computational linguistics and intelligent systems, COLINS, 366-377. (2019)
65. Frolov, V., Frolov, O., Kharchenko, V.: Classification of Diversity for Dependable and Safe Computing. In: Computational linguistics and intelligent systems, COLINS, 355-365. (2019)