Optical Character Recognition

Tetyana Zdebska, Vasyl Andrunyk, Roman Kempnyk[0000-0002-8433-976X], Vladyslav Chyhura[0000-0003-0588-1276]

Lviv Polytechnic National University, Lviv, Ukraine
kempnykrom@gmail.com

Abstract. Optical text recognition (OCR) programs are used to recognize text, that is, to translate from a graphic image into a text format. OCR (Optical character recognition) is a mechanical or electronic transfer saved handwritten, typewritten or printed text into a sequence of codes used to represent a text editor. Optical text recognition allows you to:

- edit text;
- perform search word or phrase;
- store it in a more compact form;
- show or print material without losing quality;
- analyze information;
- apply electronic translation, formatting or translation to text

Optical text recognition is a research problem in the areas of pattern recognition , artificial intelligence and computer vision. Optical text recognition systems require calibration to work with a specific font ; in earlier versions, programming required the image of each character, the program could only work with one font at a time. Currently, the most common so-called "intelligent" systems that recognize most fonts with a high degree of accuracy. Some optical text recognition systems are capable of restoring the original text formatting, including images, columns, and other non-text components.

1 Introduction

In 1929, Gustav Tauszczek received a patent for the optical text recognition method in Germany, followed by Handel, receiving a patent for his method in the United States in 1933. In 1935, Tauszczek also received a patent for the United States on his method. Tauschek's machine was a mechanical device using templates and a photodetector.

In 1950, David H. Shepard , a crypto analyst with the United States Armed Forces Security Agency, after analyzing the task of converting printed messages into a machine language for computer processing, built a machine that solves this task. After he received a US patent, he reported it to the "Washington Daily Nyuz " (27 April 1951) and "New York Times " (December 26, 1953). Shepard then founded a company that develops intelligent machines, which soon released the world's first commercial optical character recognition systems.

COLINS'2020, Volume II: Workshop. Lviv, Ukraine, April 23-24, 2020, ISSN 2523-4013
http://colins.in.ua, online

The first commercial system was installed on the "Reader's Digest" in 1955. The second system was sold «Standard Oil » read credit cards for check out. Other systems supplied by Shepard were sold in the late 1950s, including a page scanner for US National Air Force to read and transmit teletype typed messages. Later, Shepard's patent was licensed to IBM.

Approximately in 1965 "Reader's Digest" and "Ar-Si-Hey" began cooperation to create a machine readable documents using optical rozpiz navannya text intended for digitizing serial numbers of coupons " Reader's Digest" returned from advertisements. A special OCR-A font was used for the documents printed by the Ar- Si - Hey drum printer . The document reader worked directly with the RCA 301 computer (one of the first massive computers). The machine's speed was 1500 documents per minute: it checked every document except those it could not process correctly. The United States Postal Service has been using machines that use optical text recognition to sort mail since 1965, based on technologies developed by researcher Yakov Rabinov . In Europe, the first organization to use machines with optical text recognition was the British Post Office . Canada Post has been using optical character recognition systems since 1971. In the first step, the name and address of the recipient are read in the sorting center of the optical character recognition system and printed on a barcode envelope. It is applied with special ink, which is clearly visible in ultraviolet light . This is to avoid confusion with a person-filled address field that may be anywhere on the envelope.

In 1974, Ray Kurtzweil founded Kurtzweil Computer Products and began developing the first optical character recognition system capable of recognizing text printed in any type of font. Kurzweil believed that the best use of this technology was to create a reading machine for the blind, which would allow blind people to have a computer capable of reading text aloud. This device required the invention of two technologies at once - a tablet CCD scanner and a synthesizer that converts text into language. The final product was presented on January 13, 1976 during a press conference led by Kurzweil and the leaders of the National Federation of the Blind.

In 1978, Kurzweil Computer Products began selling a commercial version of a computer-based optical character recognition program. Two years later, Kurzweil sold his company Xerox, which was interested in further commercializing text recognition systems. " Kurzweil Computer Products " has become a subsidiary of "Xerox", known as " Skansoft ."

2 The Current State of Optical Text Recognition Technology

Accurate recognition of Latin characters in printed text is now possible only if clear images such as printed documents are available. Accuracy in this formulation exceeds 99%, absolute accuracy can only be achieved by subsequent human editing. The problems of recognizing handwritten "printed" text and standard handwriting, as well as printed texts of other formats (especially with very many characters) are now the subject of active research.

The accuracy of methods can be measured in several ways, and therefore can vary greatly. For example, if you encounter a specialized word that is not in the vocabulary of the related software, the error may be increased when searching for non-existent words.

Online character recognition is sometimes confused with optical character recognition. Method Optical th character recognition - of Online Booking method that works with static text form submission, while the online character recognition takes into account movements while writing. For example, on-line recognition using PenPoint OS or a tablet, you can determine which side of the line is written from left to right or left to right.

On-line handwriting recognition systems have recently become widely known as commercial products. The algorithms of such devices use the fact that the order, speed and direction of the individual sections of the input lines are known. In addition, the user can only use specific forms of writing. These methods cannot be used in software that uses scanned paper documents, so the problem of recognizing handwritten "printed" text remains open. Images with handwritten "print" text without artifacts can achieve an accuracy of 80% - 90%, but with such accuracy the image will be converted to text with dozens of errors on the page. Such technology is useful in a very limited number of applications.

Another widely researched problem is handwriting recognition . At this time, the accuracy achieved is even lower than for handwritten "printed" text. Higher scores can only be achieved by using contextual and grammatical information. For example, in the recognition process, finding entire words in the dictionary is easier than trying to parse individual characters from the text. Knowledge of grammar can also help determine whether a word is a verb or a noun. Forms of individual handwriting may sometimes not contain enough information to accurately (over 98%) recognize all handwriting.

Typically, intelligent recognition systems such as artificial neural networks are used to solve more complex recognition problems .

3 Automatic Character Recognition

After processing the document, the scanner outputs a graphic image of the document (graphic image). But the graphic is not yet a text document. It is enough for a person to look at a piece of paper with text to understand what is written on it. From a computer's point of view, the document after scanning is transformed into a set of multicolored dots, not a text document.

The problem of recognizing characters in a dot graphic is very complex. Such problems are solved with the help of special software, called image recognition.

A real technical breakthrough in this area has only taken place in recent years. Prior to this, text recognition was only possible by comparing detected point configurations with a standard sample (a standard stored in the computer's memory).

COLINS'2020, Volume II: Workshop. Lviv, Ukraine, April 23-24, 2020, ISSN 2523-4013
http://colins.in.ua, online

The authors of the programs specified the criterion of "similarity" used in the identification of characters.

Such systems were called OCR (Optical Character Recognition) and relied on specially designed fonts to facilitate this approach. If you had to deal with an arbitrary and, moreover, complex font, programs of this kind began to give serious failures.

Modern scientific advances in image recognition have literally turned the idea of optical character recognition. Modern applications can handle many different (and very fanciful) fonts without reconfiguring them. Many applications even recognize handwritten characters.

Table 3. Recognition programs

Name	*License*	*Operating Systems*	*Notes*
ABBYY FineReader	commercial proprietary	Windows; Linux , Mac OS (not for end user)	Working with different languages requires the support of the appropriate language.
Brainware [en]	Commercial [source?]	Windows	Obtaining data from documents and processing them; such as bills, messages, invoices, and billing orders
COCR2	Free	Windows 9X, ME, 2000, XP	A program for recognizing simplified and traditional Chinese characters. The main limitation of the program: for each character, the user must choose the option of recognizing it with the mouse or keyboard. But the number of recognizable characters is quite large - more than 10,000.
CuneiForm	BSD	Windows (with GUI), Linux , Mac OS, FreeBSD (CLI)	An industrial, multilingual system that can preserve text formatting and recognizes intricate arbitrary structure tables
ExperVision TypeReader & RTK	Commercial [source?]	Windows, Mac OS X, Unix , Linux , OS / 2	It was highly praised in the early 1990s.
FineReaderOnline.ru	commercial	interface: Browser	Online -OCR service that allows you to recognize multilingual text from a scanned document or photo. Converts the result to editable formats (PDF, PDF / A, DOC , RTF, XLS , TXT). Up to 10 pages a day can now be recognized for free.
FreeOCR	Apache	Interface: Browser; Server:	A platform for browser - based character recognition systems. Uses Tesseract . Large number of supported

		POSIX , Unix	languages. Project Page: FreeOCR
GOCR	GPL	Cross-platform	In the initial stage of development
HOCR	GPL	Linux	Text recognition in Hebrew
img2txt.com	commercial	interface: Browser	Online -OCR service allows you to recognize multilingual text from a scanned document or photo. Uses Tesseract
Kirtas Technologies Arabic OCR	Commercial	Windows	Can recognize Arabic and English characters on one page.
Microsoft Office Document Imaging	Commercial	Windows, Mac OS X	
Microsoft Office OneNote 2007	Commercial	Windows	
NewOCR.com	commercial	interface: Browser	Online -OCR service allows you to recognize multilingual text from a scanned document or photo. Supports 29 languages (Bulgarian, Catalan, Czech, Danish, Dutch, English, Finnish, French, German, Greek, Hungarian, Italian, Latvian, Lithuanian, Norwegian, Polish, Portuguese, Romanian, Romanian, Romanian, Polish, Portuguese, Romanian Spanish, Swedish, Tagalog, Turkish, Ukrainian, Vietnamese) and recognizes text formatted in several columns.
NovoDynamics VERUS	Commercial [source?]	?	Specializes in Middle Eastern languages
Ocrad	GPL	Unix-like , OS / 2	
OCRopus	Apache	Linux	Expandable recognition system that can use Tesseract

FineReader is an optical character recognition system developed by the Russian company ABBYY. The program for text recognition allows you to quickly and accurately convert images of documents and PDF-files into electronic format suitable for editing (Microsoft Word, Microsoft Excel, Microsoft Powerpoint , Rich Text Format , HTML, PDF / A, searchable PDF, CSV and text (plain text) files). The intuitive interface of the program allows one click to recognize documents in 189 languages. [2]

FineReader is the only OCR system in the world that operates in accordance with the principles of operation of our visual system at all stages of document processing.

Integrity (integrity) - an object as a set of parts and their spatial relationships between them. Dedication (purposefulness) - because any interpretation of the data pursues a goal, then recognition is the process of hypothesizing about the object and purposeful verification. Adaptability (adaptability) - the system stores accumulated

in the process information and use it again, that learning occurs. The corresponding technology is called IPA - in the first letters of English terms.

ABBYY has developed a special MDA algorithm (multilevel document analysis). The structure of the page is analyzed from top to bottom (from constituent elements to individual characters), and the reproduction of the electronic document after the recognition is completed from the bottom up, but at all levels, there is an additional feedback mechanism .

Most modern OCRs operate on three levels: symbols, words, pages. However, according to IPA principles, ABBYY has introduced another layer in FineReader - the entire multi-page document. First of all, it was necessary for the correct reproduction of the logical structure, which is becoming increasingly difficult in modern documents.

That's why ADRT (Adaptive Document Recognition Technology), a logical analysis and synthesis technology , was developed . Ultimately, it helps to make FineReader look as original as possible. [3]

Among the Abby FineReader-like systems are:

- SimpleOCR ;
- OmniPage ;
- Readiris ;
- CuneiForm .

We should highlight the Tesseract - a system of recognition of texts in the 1985-1994 biennium. Developed Hewlett-Packard , and since 2006 is free and licensed by Google and is developing very rapidly in recent years. [4]

4 Example of Use

Adobe Acrobat Export PDF is an online service of Adobe Document Cloud . It can easily convert PDF files to editable Word, Excel, or RTF (Rich Text Format) formats .

You cannot edit PDFs with Adobe Acrobat Export PDF. This requires the Acrobat DC application. See the Acrobat family product page .

Adobe Acrobat Export PDF supports optical text recognition (OCR) technology used when converting a PDF file to Word (.doc and .docx), Excel (.xlsx), or RTF (rich text format). Optical text recognition is the conversion of text images (scanned text) into editable characters. Recognized text supports search, correction and copy functions.

With optical text recognition enabled, Adobe Acrobat Export PDF performs optical text recognition in PDF files that contain images, vector graphics, hidden text, or a combination of these elements. (For example, Adobe Acrobat Export PDF performs optical text recognition in PDF files created from scanned documents.) Adobe Acrobat Export PDF also performs optical recognition of text that cannot be interpreted because it was incorrectly encoded in the source program.

5 Supported Languages

By default, optical text recognition uses the language specified in the About Me dialog box. Optical Text Recognition uses the selected language to recognize the scanned text. Choosing the right language enhances recognition accuracy, as the optical text recognition tool uses dictionaries for a particular language in your work. For languages that use non-Latin characters (such as Japanese), the optical text recognition tool cannot recognize and convert text unless the appropriate language is selected.

6 Conclusions

Optical text recognition is therefore a convenient tool for creating digital documents from paper originals. The textual representation allows further processing of information obtained by scanning or photographing. The relevance of text recognition has increased with the advent of e-book readers, making it easier to read. Optical recognition has made it much easier to find information in online libraries (you can find not only a volume or section of a book, but also specific sentences or words)

References

1. Optical text recognition, http://.wikipedia.org/wiki/Optical_Character Recognition
2. FineReader, http://en.wikipedia.org/wiki/ABBYY_FineReader .
3. IXBT, http://www.ixbt.com/soft/finereader.shtml .
4. Tesseract, http://en.wikipedia.org/wiki/Tesseract .
5. ABBYY FineReader Optical Character Recognition System . Version 9.0. User's Guide . 2007 ABBYY.
6. 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.
7. 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.
8. 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.
9. 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)
10. 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)
11. 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

12. 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)
13. 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)
14. 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)
15. Dovbysh, A., Alieksieiev, V.: Embedding Speech Recognition Tools for Custom Software: Engines Overview. In: Computational Linguistics and Intelligent Systems, COLINS, 2, 114-121. (2018)
16. 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 glottochronology. In: Eastern-European Journal of Enterprise Technologies, 4(2-88), 10-19. (2017)
17. 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)
18. 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)
19. 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)
20. 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)
21. 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)
22. 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)
23. 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)
24. Kapustiy, B.E., Rusyn, B.P., Tayanov, V.A.: Peculiarities of application of statistical detection criteria for problems of pattern recognition. In: Journal of Automatioin and Inrormation Science, 37(2), 30-36. (2005)
25. 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)
26. 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 Mashiny, 2, 8-12. (2005)
27. 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)
28. Rusyn, B., Torska, R., Kobasyar, M.: Application of the cellular automata for obtaining pitting images during simulation process of thei growth. In: Advances in Intelligent Systems and Computing, 242, 299-306. (2014)
29. 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)
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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 – Стор.43-56.
36. 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.
37. 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–RZESZУW,2016. – pp. 111-125. - ISSN 2084-5715. - http://www.pan-ol.lublin.pl/wydawnictwa/Econtechmod.html.

38. 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.
39. 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, nomber 2. – P. 85–94. – Bibliography: 60 titles.
40. 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.
41. 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.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.
47. Lytvyn Vasyl. The Contextual Search Method Based on Domain Thesaurus / Vasyl 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.

48. Kanishcheva Olga. Method of Integration and Content Management of the Information Resources Network / Olga Kanishcheva, Victoria Vysotska, Lyubomyr Chyrun, Aleksandr Gozhyj // Advances in Intelligent Systems and Computing. – Vol. 689. – ISSN 2194-5357 ISSN 2194-5365 (electronic). - Springer International Publishing AG 2017. –PP. 204-216.
49. Jun Su. Information resources processing using linguistic analysis of textual content / Jun Su, Victoria Vysotska, Anatoliy Sachenko, Vasyl 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.
50. 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.
51. 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)
52. Zhezhnych, P., Markiv, O.: Recognition of tourism documentation fragments from webpage posts. In: 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET, 948-951. (2018)
53. 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)
54. 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)
55. Shanidze, O., Petrasova, S.: Extraction of semantic relations from Wikipedia text corpus. In: Computational linguistics and intelligent systems, COLINS, 2, 74-75. (2019)
56. Razno, M.: Machine learning text classification model with NLP approach. In: Computational linguistics and intelligent systems, COLINS, 2, 71-73. (2019)
57. Manuilov, I., Petrasova, S.: Method for paraphrase extractionfrom the news text corpus. In: Computational linguistics and intelligent systems, COLINS, 2, 69-70. (2019)
58. Lytvynenko, J.: Identify of the substantive, attribute, and verb collocations in russian text. In: Computational linguistics and intelligent systems, COLINS, 2, 66-68. (2019)
59. Liutenko, I., Kurasov, O.: Consideration of the software tests quality evaluation problem. In: Computational linguistics and intelligent systems, COLINS, 2, 62-65. (2019)
60. Drobot, T.: Data-to-text generation for domain-specific purposes. In: Computational linguistics and intelligent systems, COLINS, 2, 60-61. (2019)
61. Chuiko, V., Khairova, N.: Semantic similarity identification for short text fragments. In: Computational linguistics and intelligent systems, COLINS, 2, 57-59. (2019)
62. 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)
63. 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)