Defining Author's Style for Plagiarism Detection in Academic Environment

Victoria Vysotska
Information Systems and Network Department
Lviv Polytechnic National University
Lviv, Ukraine
Victoria.A.Vysotska@lpnu.ua

Vasyl Lytvyn
Information Systems and Network Department
Lviv Polytechnic National University
Lviv, Ukraine
Vasyl.V.Lytvyn@lpnu.ua

Yevhen Burov
Information Systems and Network Department
Lviv Polytechnic National University
Lviv, Ukraine
Yevhen.V.Burov@lpnu.ua

Andriy Demchuk
Information Systems and Network Department
Lviv Polytechnic National University
Lviv, Ukraine
Andrii.B.Demchuk@lpnu.ua

Abstract—The usage of linguometry and stylometry technologies for author's style detection is discussed. The statistical linguistic analysis of author's text uses the advantages of content monitoring based on NLP methods for stop words definition. They are used in stylometry for definition of attribution degree of analyzed text to the specific author. The formal approach for Ukrainian language text author's style definition is proposed.

Keywords—plagiarism, plagiarism detection, author's style, statystic linguistic analysis, quantitative linguistics, authorship attribution.

I. INTRODUCTION

The strategic task of scientific and educational organizations is the formation of scientific elite, able to contribute to economic growth by introducing innovative products in industry, agriculture, medical services and information technologies. They would be able to support the sustainable and continuous growth of Ukrainian economy. The effectiveness of education and science is a constant concern for government and public organizations. Plagiarism negatively affects the quality of education and science [1]. Academic integrity is a rather broad term. It is understood as a set of ethical principles and rules defined by laws which should be followed by all participants of educational process in study, scientific research with the ultimate purpose of creating trust in objectivity of educational or scientific accomplishments. Attaining academic integrity requires a systemic approach combining the implementation of various organizational, educational, and technical measures. Among important technical solutions in this area is the creation of repository of academic texts of organization and National repository of academic texts (NRAT); and using information system (IS) for plagiarism detection in scientific articles which detects similarities based on data from NRAT and Internet. In article [2] plagiarism is subdivided in four kinds, each of them having the distinct purpose. Depending of activity type and application area, plagiarism can be:

- professional (acquisition of intellectual, creative and professional achievements of other people with a professional purpose;

- educational and scientific (appropriation of intellectual property uniquely in process of attaining a scientific degree, or qualification certification, or confirming existing degree);

- social (similar to professional, but not occurring in professional environment);

- normative (appropriation of results of methodical, scientific work, laws and bills developed by other people, also practical solutions).

Depending on form, we establish following types of plagiarism:

- Full or partial copying of author's work using linguistic, lexical and technological interpretation (implemented by most of existing plagiarism detectors).

- Appropriation of the main idea (hard to detect, because there are no methods in existing systems allowing to extract and analyze meaning of document).

- Plagiarism implying the use of references, while citing other works, referencing non-existing or non-relevant sources, presenting facts from other research without referencing it (partially implemented in some plagiarism detecting systems);

- When plagiarized work is not created by putative author himself, but bought from certain organization or person producing such works commercially and attributed to buyer. The detection of such type of plagiarism requires the definition of original writer's style using as a sample his other works. This important problem is addressed in this article.

II. BACKGROUND ANALYSIS

For automatic detection of language formatted excerpts of text are analyzed: letters are ordered in diminishing order according to frequencies of their appearance in excerpt (frequencies are shown), capitals and small letters are not distinguished. We can analyze the data and detect the

128

language of text using one of three methods presented in [3-9]:

- Frequencies of vowels and consonants in text,

- Frequencies of voiced, voiceless, nasal and other types of letters and their combinations,

- Frequencies of letter's usage.

In order to research peculiarities of author's style various individualized lexical metrics were defined. Among them are continuity, lexical variety, syntactic complexity, indices of concentration and uniqueness for fragments from author's text and other text. Next, using those metrics, internal dynamics of text is analyzed and finally, the degree of appropriation of analyzed text to a specific author is defined [10]. In order to define the degree of plagiarism a summary grouping table is built. In this table are placed, calculated previously for sets of texts with similar content, the average group values of continuity, lexical variety, syntactic complexity, indices of concentration and uniqueness [11].

The area of standard deviations is defined and thus the lexical similarity of analyzed fragments is compared to benchmark [12]. The author of text is defined based on analysis of his formatted text [13]. Words are placed in falling order according to their usage frequency in text fragment. The type of speech (direct or indirect) is noted. Common names are removed from text fragment. Author of text fragment or of a whole text is defined when possible using frequency dictionaries [14]. The analysis of authorship is based of definition of differences between individual author styles [15]. The uniqueness of style make author's language dynamic, memorable, easier to understand. However it is important to differentiate between individual style characteristics and characteristics common to authors [16]. The degree of authorship attribution of particular work includes credibility, authenticity of such work characteristics as author, time and place of creation. This degree is defined based on analysis of stylistic and technical distinctions [17].

III. STATEMENT OF PROBLEM

Our work aims to detect plagiarism in scientific articles by analysis of author style on a deeper level. We are treating this task as a text classification problem (fig. 1), where every document belongs only to a specific category/class/author. As initial dataset we use one-author articles of 100 Ukrainian scientists, 10 articles from each.

Thus, we analyze the collection of 1000 texts. For each category benchmarks which correspond to the style are defined. For each of researched text the coefficient of attribution to specific categories is calculated. For author's style definition the methods of machine learning were used (Bayesian classificator and reference vector method). Using our dataset we also obtained such quantitative characteristics of author's style as coefficient of text variability, degree of syntactic complexity, continuity, and uniqueness and concentration indices [3-5].

The stage of text preprocessing is rather time consuming and consists of primary processing of linguistic data (building of distribution sequences, calculating statistics and other linguometric characteristics), lexicographical processing of textual data (creating frequental and alphabetically frequental dictionaries for texts of specific

author, dictionaries of concordances, referential words, and author's style keywords) and applying methods of stylometry for authorship attribution [18-24].
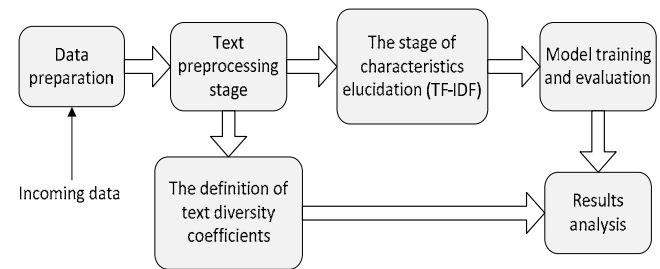

Fig. 1.   The general scheme of authorhip attribution workflow

It is practically impossible to do this without preceding analysis of author's works represented in form of benchmark. The linguistic analysis and interpretation of stylistic properties and regularities of author's style is done using methods of content analysis and has following steps: the correct selection of texts; lemmatization of text units; removing heterogeneity of text units; building dictionaries and based on them creating statistical distributions within corresponding frequential linguistic dictionary scales; search of parameters which adequately represent the structure of frequential dictionary; reviewing parameters for effectiveness; mathematical modeling and analysis of lexical and statistical distributions; creating statistical classifications; results interpretation and definition of benchmarked coefficients of text diversity for the specific author. An important part of this process is the definition of correct keyword dictionaries of author's style.

We developed a system allowing to select language(s) used in text. The access to the process of keywords definition based of thematic words stems is provided by Victana (victana.lviv.ua) resource [25]. When building such dictionaries from Ukrainian texts the word stems should be taken without inflections.

IV. THE METHOD OF AUTHOR'S STYLE DEFINITION FOR TEXTUAL CONTENT

The linguistic and statistical basis for research of text attribution is based on works [3, 18–23]:

1) primary processing of linguistic data (building distribution sequences, calculating statistics, statistical metrics and other linguometric parameters),

2) lexicographic processing of textual data (creation of frequential and alphabetically frequential dictionaries, dictionaries of concordances, referential words, and author's style keywords).

The usage of linguometric methods for statistical description of text supports research in the domain of authorship [24]. The method of linguistic analysis and interpretation of stylistic features and regularities of writer's style (or style of specific historical period) uses algorithm 1.

Algorithm 1. The linguistic analysis and interpretation of stylistic features and regularities of specific author's style.

Stage 1. Texts selection. It is important how texts are selected and the size of selection: for reliable definition of characteristics the required selection size is at least 18 thousand words [26-31].

129

Stage 2. The lemmatization of textual units. Unification of different words according to lemmas of language [5].

Stage 3. Removing disparity of textual units. Resolving the problem of text units' disparity, such as the usage of direct and indirect speeches.

Stage 4.Building a system of frequential dictionaries and, based on it, creation of the statistical distributions within corresponding frequency scales. Frequential dictionary is a type of dictionary where the usage frequency of specific language object is noted (word composition, word form, collocations, idioms, phrases) in different texts of specific size. Usually the absolute and relative frequency of language units usage is shown, in fall down order [3].

Stage 5. Search for parameters which adequately reflect the structure of frequential dictionary. The number of parameters is different. For example, for description of French texts from XVII century 51 parameters were proposed [26-31]. The common parameters found in [27–30] allow to formulate several linguo-statistical methods for text research:

– the method of supporting words (calculation of usage frequency and percent of auxiliary words [18–22]: such as pronouns, prepositions, conjunctions and determiners;

– the method of punctuation marks (only calculate the number of internal and external punctuation marks);

– the method of words (only calculate the number of words of specified length);

– the method of sentences (only calculate the number of sentences of specific length);

– the syntactic method (calculate the number of punctuation marks, words and sentences of specified length);

– the combined method (the combination of supporting words method and syntactic method).

Stage 6. Checking the effectiveness of parameters. Applying known methods for selected parameters effectiveness verification.

Stage 7. Mathematical modeling of lexico-statystical distributions. Applying known methods of mathematical modeling of lexico-statystical distribution.

Stage 8. Building of statistical classifications (author's benchmarks) reflecting the stylistic regularities within works of specified author or specific literary epoch (or the sequence of literary epochs).

Stage 9. The interpretation of obtained results from the perspective of historical and linguistic knowledge, general and historical stylistics.

Using algorithm 1 we perform the task of authorship attribution which can be described as follows. Suppose, that there is a benchmark statistically reflecting all author's works. We should evaluate the belonging of specific texts to benchmark using corresponding methods. Let's consider as example the works of Author 1 and her publications from [32-40]. Assume that author's benchmark is already built, that tasks of selecting texts, lemmatization and non-uniformness problems are already resolved and frequential dictionary is created [3, 41-48]. We will use the method of supporting words for attribution and will represent results as

correlation coefficients and also graphically. Separately, we will also note the evolution of relevance of one of text's parameters – auxiliary words- for authorship attribution: pronouns, conjunctions, prepositions [49-64].

For individual writer's style definition the auxiliary words are significant because they are not influenced by book theme or content [3]. We will assume that this parameter is effective for authorship attribution and will use a list of 71 auxiliary stop-words [41-48, 55-69].

V. THE RESULTS OF EXPERIMENTS

Let's take four arbitrary fragments from [3-5, 25], formatted depending on attribution method: from each fragment only prepositions, conjunctions and pronouns are used. The overall number of word usages is calculated and common names are excluded. For each of fragments we define the absolute frequency (AF) and relative frequency (RF) of auxiliary word appearance and also the relative frequency of this word appearance in benchmark. Figure 2 shows the graphical representation of relative frequency of stop words appearance in Fragment 1 and in benchmark. The coefficient of correlation for auxiliary words in this case is $R_{e-y1}=0,6076$.
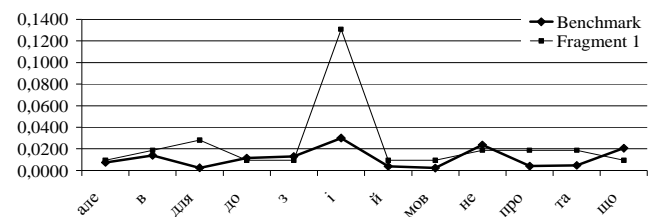

Fig. 2. Relative frequency of auxiliary words appearance in fragment 1 and benchmark
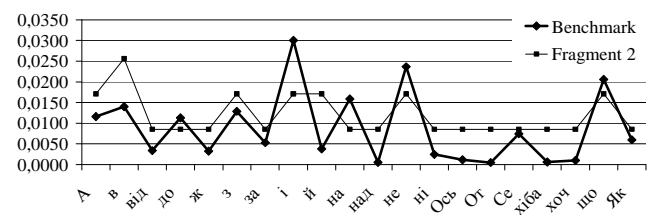

Fig. 3. Relative frequency of auxiliary words appearance in fragment 2 and benchmark.
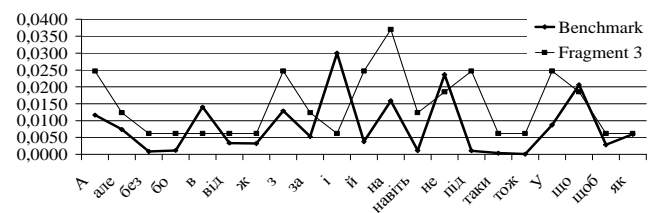

Fig. 4. Relative frequency of auxiliary words appearance in fragment 3 and benchmark
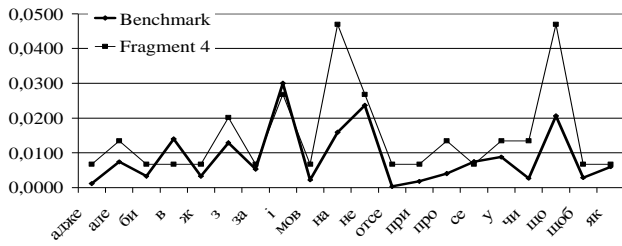
130

Fig. 5. Relative frequency of auxiliary words appearance in fragment 4 and benchmark

The graphical representation of relative frequency of auxiliary words appearance in fragment 2 and benchmark is shown on fig. 3. The coefficient of correlation in this case is $R_{e-Y2}=0,7066$. The graphical representation of relative frequency of auxiliary words appearance in fragment 3 and in benchmark is shown on fig. 4. The coefficient of correlation in this case is $R_{e-Y3}=0,2810$. The relative frequencies of auxiliary words in fragment 4 and benchmark are presented in fig.5. The coefficient of correlation is $R_{e-Y4}=0,7326$. The correlation coefficients for each auxiliary word for fragments 1-4 are summarized in Table I. After the analysis of correlation coefficients for auxiliary words we conclude that probability of relatedness of fragments to benchmark is largest for fragment 4, the next is fragment 2 and then fragments 1 and 3. Let's note, that for all four fragments there are high correlations for prepositions. This could be interpreted as prepositions not influencing the author's style. Additionally, for selected text fragments, we analyzed the frequency of only pronouns and only conjunctions appearance and calculated corresponding correlation coefficients. The comparison of results is shown in Table II.

TABLE I. CORRELATION COEFFECIENTS FOR AUXILIARY WORDS

Fragment	Pronoun	Conjunction	Preposition
1	$R_{e-Y1Z}=0,72$	$R_{e-Y1S}=0,79$	$R_{e-Y1C}=1$
2	$R_{e-Y2Z}=0,4928$	$R_{e-Y2S}=0,5714$	$R_{e-Y2C}=0,9580$
3	$R_{e-Y3Z}=0,1517$	$R_{e-Y3S}=0,1624$	$R_{e-Y3C}=0,8800$
4	$R_{e-Y4Z}=0,5639$	$R_{e-Y4S}=0,9544$	$R_{e-Y4C}=0,9594$

TABLE II. CORRELATION COEFICIENTS FOR EVERY FRAGMENT

Coefficient	Fragment 1	Fragment 2	Fragment 3	Fragment 4
R_{e-y}	0,6076	0,7066	0,2810	0,7326
R'_{e-y}	0,6900	0,4913	0,2254	0,6905

Fragment 4 is still has the biggest probability of being related to benchmark, after it with a small gap go fragments 1, 2, 3. The order is the same as was in previous experiment. For confirmation of results we refer to [25], which provided fragments for experiment. Thus, the usage of supporting words method lead to following results: among researched fragments the greatest probability of relatedness to benchmark is obtained for the fragment which in fact belongs to it [25]. Other results also confirm the applicability of supporting words method for texts authorship attribution. Accordingly, in first experiment the second greater probability of being related to benchmark is assigned to fragment of another text by the same author. Fragment 1 which also belongs to benchmark, lost to fragment 4 only one tenth in correlation coefficient value. Also adequate is the result for fragment 3, which is represents a text nearly

one hundred years distant in its creation time from benchmark. The hypothesis made in [25] that prepositions as method parameters have no influence on author's style definition lead to reduction of correlation coefficients, but put the probability of relatedness to benchmark in right order. Most of all, the difference between correlation coefficients for fragments 1 and 4 was considerably reduced and had value of 0,0005. However, for confirmation or refutation of claim that prepositions are not an important factor in authorship definition additional research should be done.

VI. CONCLUSION

In this work the importance of textual works analysis for plagiarism was shown, types of plagiarism were reviewed and approach to detect author's style was described. The authors researched machine learning methods and linguistic metrics for definition of author's style based on corpus of Ukrainian academic articles. The publications authored by several people present a separate problem. The definition of style in this case is more complex because styles of different authors are superimposing. In this case it is difficult to detect whether the academic work was created by commercial writer.

REFERENCES

[1] L. Luparenko, "Instrumentariy vyyavlennya plahiatu v naukovykh robotakh: analiz prohramnykh rishen," Informatsiyni tekhnolohiyi i zasoby navchannya, Vol. 2(40), 2014, pp. 151-169.
[2] V. Petrenko, "Ponyattya ta vydy plahiatu," Chasopys tsyvilistyky, Vol. 14, 2013, pp. 128-131.
[3] V. Lytvyn, V. Vysotska, P. Pukach, I. Bobyk, and D. Uhryn, "Development of a method for the recognition of author's style in the Ukrainian language texts based on linguometry, stylemetry and glottochronology," Eastern-European Journal of Enterprise Technologies, vol. 4/2(88), pp. 10–18, 2017.
[4] V. Lytvyn, V. Vysotska, P. Pukach, O. Brodyak, and D. Ugryn, "Development of a method for determining the keywords in the slavic language texts based on the technology of web mining," EasternEuropean Journal of Enterprise Technologies, vol. 2/2(86), pp. 4–12, 2017.
[5] V. Lytvyn, P. Pukach, I. Bobyk, and V. Vysotska, "The method of formation of the status of personality understanding based on the content analysis," Eastern–European Journal of Enterprise Technologies, vol.5/2(83), pp. 4–12, 2016.
[6] B. Mobasher, Data mining for web personalization," The adaptive web, Springer Berlin Heidelberg, 2007, pp. 90–135.
[7] C. E. Dinuca, and D. Ciobanu, "Web Content Mining," University of Petrosani, Economics, pp. 85-92, 2012.
[8] G. Xu, Y. Zhang, and L. Li. Web content mining. Web Mining and Social Networking. Springer US, 2011, pp. 71–87.
[9] I. Khomytska, and V. Teslyuk, "The Method of Statistical Analysis of the Scientific, Colloquial, Belles-Lettres and Newspaper Styles on the Phonological Level," Advances in Intelligent Systems and Computing, pp. 149–163, 2017.
[10] I. Khomytska, and V. Teslyuk, "Specifics of Phonostatistical Structure of the Scientific Style in English Style System," Computer Science and Information Technologies, Proc. of the XI–th Int. Conf. CSIT, pp. 129–131, 2016.
[11] I. Khomytska, V. Teslyuk, "Modelling of Phonostatistical Structures of English Backlingual Phoneme Group in Style System", Proceedings of CADMS, pp. 324-327, 2017.
[12] I. Khomytska, V. Teslyuk, "Modelling of Phonostatistical Structures of the Colloquial and Newspaper Styles in English Sonorant Phoneme Group", Computer Science and Information Technologies, CSIT, pp. 67-70, 2017.
[13] A. Anisimov, and A. Marchenko, "Sistema obrabotki tekstov na yestestvennom yazyke," Iskusstvennyy intellekt, 4, 157–163, 2002.
[14] V. Perebiynis, "Matematychna lingvistyka," Ukrains'ka yentsiklopediya, pp. 287-302, 2000.
[15] S. Buk, "Osnovy statystychnoy lingvistyky," I. Franka LNU, 2008.
[16] M. Kochergan, "Vstup do movoznavstva," Akademíya, 2005.

131

[17] V. Perebiynis, Statystychni metody dlya lingvistiv. Nova kniga, 2013.

[18] M.A. Gonçalves, "Streams, Structures, Spaces, Scenarios, and Societies (5S): A Formal Model for Digital Library Framework and Its Applications", PhD thesis, 2004.

[19] A. Varfolomeyev, "Psikhosemantika slova i lingvostatistika teksta," Kaliningrad, KGU, 2000.

[20] R. Manekin, "Kognitivnaya stilometriya: k postanovke problemy", (http://www.manekin.narod.ru/hist/styl.htm).

[21] B. Rusyn, O. Lutsyk, O. Lysak, A. Lukeniuk, and L. Pohreliuk, "Lossless Image Compression in the Remote Sensing Applications", Data stream mining & processing, DSMP, pp.195-198, 2016.

[22] Kowalik Dagmara, "Polish vocational competence standards for the needs of audult education and the Europen labour marke", International Conference on Advanced Information Engineering and Education Science, ICAEES, pp. 95-98, 2013.

[23] R. Tkachenko, P. Tkachenko, I. Izonin, Y. Tsymbal, "Learning-based image scaling using neural-like structure of geometric transformation paradigm", In Studies in Computational Intelligence, vol. 730, Springer Verlag, pp. 537–565,. 2018.

[24] Y. Rashkevych, D. Peleshko, O. Vynokurova, I. Izonin and N. Lotoshynska, "Single-frame image super-resolution based on singular square matrix operator," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 944-948, 2017.

[25] Victana, (http://victana.lviv.ua/index.php/kliuchovi–slova).

[26] O. Kanishcheva, V. Vysotska, L. Chyrun, and A. Gozhyj, "Method of Integration and Content Management of the Information Resources Network," Advances in Intelligent Systems and Computing, Springer, 2017, pp. 204–216.

[27] J. Su, V. Vysotska, A. Sachenko, V. Lytvyn, and Y. Burov, "Information resources processing using linguistic analysis of textual content," Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 9th IEEE International Conference, Bucharest, Romania, pp. 573–578, 2017.

[28] V. Lytvyn, V. Vysotska, O. Veres, I. Rishnyak, and H. Rishnyak, "The Risk Management Modelling in Multi Project Environment," Computer Science and Information Technologies, XII-th Int. Conf. CSIT, pp. 32–35, 2017.

[29] M. Korobchinsky, V. Vysotska, L. Chyrun, and L. Chyrun, "Peculiarities of Content Forming and Analysis in Internet Newspaper Covering Music News." Computer Science and Information Technologies, XII-th Int. Conf. CSIT, pp. 52–57, 2017.

[30] O. Naum, L. Chyrun, O. Kanishcheva, and V. Vysotska, "Intellectual System Design for Content Formation," Computer Science and Information Technologies, XII-th Int. Conf. CSIT, pp. 131–138, 2017.

[31] V. Lytvyn, V. Vysotska, Y. Burov, O. Veres, and I. Rishnyak, "The Contextual Search Method Based on Domain Thesaurus," Advances in Intelligent Systems and Computing, pp. 310–319, 2017.

[32] M. Davydov, and O. Lozynska, "Information System for Translation into Ukrainian Sign Language on Mobile Devices," Computer Science and Information Technologies, CSIT, pp. 48-51, 2017.

[33] Jivani Anjali Ganesh. "A Comparative Study of Stemming Algorithms," Int. J. Comp. Tech. Appl., vol. 2(6), 1930–1938, 2011.

[34] A. Mishler, E.S. Crabb, S. Paletz, B. Hefright, and E. Golonka, "Using Structural Topic Modeling to Detect Events and Cluster Twitter Users in the Ukrainian Crisis," Communications in Computer and Information Science, vol. 528. Springer, pp. 639–644, 2015.

[35] M. Davydov, and O. Lozynska, "Mathematical Method of Translation into Ukrainian Sign Language Based on Ontologies," Advances in Intelligent Systems and Computing, vol. 689, pp. 89-100, 2018.

[36] O. Chernukha, and Y. Bilushchak, "Mathematical modeling of random concentration field and its second moments in a semispace with erlangian disrtibution of layered inclusions", Task Quarterly, vol. 20(3), pp. 295-334, 2016.

[37] K. Kowalska, Cai Di, and S. Wade, "Sentiment analysis of polish texts," Inter. J. of Computer and Communication Engineering, vol. 1.1, pp. 39–41, 2012.

[38] N. Kotsyba, "The current state of work on the Polish-Ukrainian Parallel Corpus," Organization and Development of Digital Lexical Resources, pp. 55–60, 2009.

[39] M. Davydov, and O. Lozynska, "Linguistic Models of Assistive Computer Technologies for Cognition and Communication," Computer Science and Information Technologies, XI-th Int. Conf. CSIT, pp. 171-175, 2017.

[40] V. Lytvyn, V. Vysotska, O. Veres, I. Rishnyak, and H. Rishnyak, "Classification Methods of Text Documents Using Ontology Based Approach," Advances in Intelligent Systems and Computing, Springer, 512, pp. 229–240, 2017.

[41] V. Vysotska, "Linguistic Analysis of Textual Commercial Content for Information Resources Processing," Modern Problems of Radio Engineering, Telecommunications and Computer Science, TCSET, pp. 709–713, 2016.

[42] V. Vysotska, L. Chyrun, and L. Chyrun, "Information Technology of Processing Information Resources in Electronic Content Commerce Systems," Computer Science and Information Technologies, CSIT, pp. 212–222, 2016.

[43] V. Vysotska, L. Chyrun, and L. Chyrun, "The Commercial Content Digest Formation and Distributional Process," Computer Science and Information Technologies, XI–th Int. Conf. CSIT, pp. 186–189, 2016.

[44] V. Lytvyn, V. Vysotska, O. Veres, I. Rishnyak, and H. Rishnyak, "Content linguistic analysis methods for textual documents classification," Computer Science and Information Technologies, CSIT, pp. 190–192, 2016.

[45] V. Lytvyn, and V. Vysotska, "Designing architecture of electronic content commerce system," Computer Science and Information Technologies, X–th Int. Conf. CSIT, 115–119, 2015.

[46] V. Vysotska, and L. Chyrun, "Analysis features of information resources processing," Computer Science and Information Technologies, pp. 124–128, 2015.

[47] Vasyl Lytvyn, Victoria Vysotska, Dmytro Dosyn, Roman Holoschuk, and Zoriana Rybchak, "Application of Sentence Parsing for Determining Keywords in Ukrainian Texts," Computer Science and Information Technologies, pp. 326–331, 2017.

[48] O. Maksymiv, T. Rak, and D. Peleshko, "Video-based Flame Detection using LBP-based Descriptor: Influences of Classifiers Variety on Detection Efficiency,". International Journal of Intelligent Systems and Applications, vol. 9(2), pp. 42–48, 2017.

[49] D. Peleshko, T. Rak, and I. Izonin, "Image Superresolution via Divergence Matrix and Automatic Detection of Crossover," International Journal of Intelligent Systems and Application, vol. 8(12), pp. 1–8, 2016.

[50] O. Bazylyk, P. Taradaha, O. Nadobko, L. Chyrun, and T. Shestakevych, "The results of software complex OPTAN use for modeling and optimization of standard engineering processes of printed circuit boards manufacturing," Proceedings of the 11th International Conference TCSET, pp. 107–108, 2012.

[51] A. Bondariev, M. Kiselychnyk, O. Nadobko, L. Nedostup, L. Chyrun, and T. Shestakevych, "The software complex development for modeling and optimizing of processes of radio-engineering equipment quality providing at the stage of manufacture," Proceedings of the 11th International Conference TCSET, pp. 159, 2012.

[52] V. Riznyk, "Multi-modular Optimum Coding Systems Based on Remarkable Geometric Properties of Space," Advances in Intelligent Systems and Computing, 512, pp. 129–148, 2017.

[53] V. Teslyuk, V. Beregovskyi, P. Denysyuk, T. Teslyuk, and A. Lozynskyi, "Development and Implementation of the Technical Accident Prevention Subsystem for the Smart Home System," International Journal of Intelligent Systems and Applications, vol. 10(1), pp. 1–8, 2018.

[54] T. Basyuk, "The main reasons of attendance falling of internet resource," CSIT, pp. 91–93, 2015.

[55] V. Pasichnyk, and T. Shestakevych, "The model of data analysis of the psychophysiological survey results," Advances in Intelligent Systems and Computing, vol. 512, pp. 271–281, 2017.

[56] P. Zhezhnych, and O. Markiv, "Linguistic Comparison Quality Evaluation of Web-Site Content with Tourism Documentation Objects," Advances in Intelligent Systems and Computing, vol. 689, pp. 656–667, 2018.

[57] P. Kravets, "The game method for orthonormal systems construction", The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2007.

[58] P. Kravets, "Game Model of Dragonfly Animat Self-Learning", Perspective Technologies and Methods in MEMS Design, pp. 195-201, 2016.

[59] P. Kravets, "The control agent with fuzzy logic", Perspective Technologies and Methods in MEMS Design, pp. 40-41, 2010.

[60] Google – word2vec, (github.com/danielfrg/word2vec/blob/master/examples/word2vec.ipynb).

[61] V. Lytvyn, V. Vysotska, P. Pukach, M. Vovk, and D. Ugryn, "Method of functioning of intelligent agents, designed to solve action planning problems based on ontological approach", Eastern-European Journal of Enterprise Technologies, vol. 3/2(87), pp. 11-17. 2017.

[62] V. Lytvyn, V. Vysotska, L. Chyrun, and L. Chyrun, "Distance Learning Method for Modern Youth Promotion and Involvement in

Independent Scientific Researches", Data Stream Mining & Processing, pp. 269-274, 2016.

[63] V. Vysotska, I, Rishnyak, and L. Chyrun, "Analysis and evaluation of risks in electronic commerce", CAD Systems in Microelectronics, 9th International Conference, pp. 332-333, 2007.

[64] V. Lytvyn, V. Vysotska, O. Veres, I. Rishnyak, and H. Rishnyak, "The Risk Management Modelling in Multi Project Environment",. Computer Science and Information Technologies, pp. 32-35, 2017.

[65] M.A. Gonçalves, E.A. Fox, L.T.nWatson, and N.A. Kipp, "Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries", ACM Transactions on Information Systems (TOIS), vol. 22(2), pp. 270-312, 2004.

[66] A. Pérez, M. Enrech, "Virtual Library Services for a Virtual University: User-Oriented Virtual Sites in an Open Library", EADTU, Paris, 1999.

[67] A. Pérez, M Enrech, "Defining library services for a virtual community", Libraries Without Walls Conference, Lesvos, Grecia, Centre for research in Library and Information Management, 1999.

[68] V. Pasichnyk, T. Shestakevych, "The method of education format ascertaining in program system of inclusive education support", Computer Science and Information Technologies, 279-284, 2017.

[69] V. Pasichnyk, T. Shestakevych, "The application of multivariative data analysis technology to support inclusive education", Computer Science and Information Technologies, pp. 88-90, 2015.