

ICEST 2020
**International Conference on Economic and Social Trends for Sustainability of
Modern Society**

**SOFTWARE FOR SENSE COMPATIBILITY ANALYSIS OF
EDUCATIONAL TEXTS**

G. R. Rybakova (a)*, A. Yu. Andreeva (b), I. V. Krotova (c), T. L. Kamoza (d)

*Corresponding author

(a) Siberian Federal University, 79 Svobodny Ave., Krasnoyarsk, Russia, rbkv@yandex.ru,

(b) Altai State Technical University, Lenin Ave., 46, Barnaul, Russia, ang_reg@mail.ru,

(c) Siberian Federal University, 79 Svobodny Ave., Krasnoyarsk, Russia, irakrotova@inbox.ru,

(d) Siberian Federal University, 79 Svobodny Ave., Krasnoyarsk, Russia, tat.kamoza@yandex.ru

Abstract

One of the areas of innovative development in pedagogy is the search for ways to increase the availability of educational text material due to a decrease in the reading competence of students. To overcome the fragmented perception of educational texts it is necessary to draw on the results of modern research in various sciences. A preliminary analysis showed that the effectiveness of the perception and assimilation of information can be increased by involving the unconscious component of the psyche by fixing attention on the most repeated terms and concepts. One of the solutions is the development of a mathematical model and software product for didactic analysis of the compatibility of educational information by defining an expanded thesaurus, which in relation to educational information implies terminology as the main fixer of the meanings of educational information. The software was developed based on the Django framework, the mathematical model is implemented in python 3.7 using the scikit-learn library. The developed tool is relevant in the context of lifelong education for the analysis and processing of educational texts without taking into account their semantic compatibility and patterns of perception of educational information. The gradual increment of the thesaurus within the framework of the necessary level of perception will allow us to level negative trends in changing the quality of perception of information by students.

2357-1330 © 2020 Published by European Publisher.

Keywords: Educational information, subject thesaurus, semantic compatibility of educational texts, software, mathematical model.



This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Modern challenges dictate the need for the formation of didactic analysis tools for the sources of educational information, which include the classic traditional textbook in its paper embodiment, as well as electronic and distance learning resources. If the educational literature for school education has always been under the scrutiny of methodologists, the educational literature of other levels of education is more often required only to match the content with the current level of development of the corresponding field of knowledge. At the same time, disputes about the quality of school textbooks and the ambiguity of approaches to the formation of their substantive part do not cease.

The amount of meanings embedded in the content of educational information can be determined from different perspectives. But if we are talking about academic disciplines representing independent branches of knowledge, their specificity is maximally reflected in the terminological component. When studying interdisciplinary issues, a different interpretation of concepts may arise with the addition of a semantic load of terms (Klochkov et al., 2008; Novikov, 2019; Popova, 2012). In this case, for mastering the terminology, a smooth transition to different levels of complexity is important (Metcalf, 2017; LeCun et al., 2015; Wilson et al., 2019 and others). To reduce the influence of material of factors associated with an objective decrease in the level of reading competence on the assimilation (Feldstein, 2013b) and the ability to perceive the meanings of large amounts of textual information, it is necessary to use an unconscious component of the students' psyche, taking into account modern interdisciplinary trends in the field of its study (Chernigovskaya, 2015; Chernigovskaya et al., 2016; Damasio, 2018; Dean, 2018; Filippova, 2006; Klementovich et al., 2016; Klochkov et al., 2008; Kryukova et al., 2017; Morozov & Spiridonov, 2019; Pervushina & Osetrin, 2017; Verbitskaya, 2019). Such an opportunity arises thanks to the mechanism of gradual increment of the special thesaurus of disciplines, which is able to form the continuity or compatibility of educational texts among themselves. To realize this possibility, we need a tool that allows analyzing information from a set of educational texts (collections of study books, a sequence of sections, disciplines, wordings), which is quite feasible using computer processing methods.

2. Problem Statement

The main research issues are formed by the following aspects:

- the need to involve the unconscious component of the psyche in the process of perceiving educational information;
- the choice of a method for analyzing the transfer of meanings embedded in an array of educational information;
- development of software for didactic analysis of semantic compatibility of texts

The main task in this context is to attract the unconscious component of the psyche in the perception of information by means of a systematic increment of the thesaurus, which will ensure text compatibility. The development of software based on this task gives the tool that will allow to analyze texts for semantic compatibility and highlight the most significant terms by the frequency of their presence in the texts to correct them subsequently.

2.1. The need to involve the unconscious component of the psyche in the process of perceiving educational information

The urgency of effective perception of educational information is not lost over the years. It undergoes changes due to objective social, technical, biological reasons including getting new ideas about the process of perception, the role of the unconscious, the characteristics of the brain, the fixation of which is given the works of outstanding teachers, psychologists, representatives of neurobiological and some other sciences (Anokhin & Velichkovsky, 2011; Chernigovskaya, 2010; Chernigovskaya et al., 2016; Damasio, 2018; Dean, 2018; Feldstein, 2013a, etc.).

In the works of Feldstein (2013a, 2013b), a generalization of the results of a large volume of research made it possible to identify the main problems associated with the ability to perceive and assimilate educational information regarding the latest generations of students. The most significant problems include a decrease in readership, manifested in the desire of pupils and students to small text forms, in the inability to holistic perception of the common semantic boundaries of larger amounts of information. All this happens against the background of a reduced motivation for learning in general, to a decrease in the authority of a teacher, mentor, and parent. The reason for this situation can be very simple including possibilities of obtaining and the availability of any information at the modern level of development of information technology. However, in practice, this leads to the loss of the time resource at different stages of development and formation of the child's brain, when the replacement of traditional development methods occurs at the expense of technical means with the so-called "smart interface". One of their negative factors of their use, as shown in the work of Morozova and Novikova (2013), is not only the tension of the work of the organs of vision due to the specific adjustment of the visual apparatus to the pixel image of the screens, but also the involvement of a large number of different areas of the brain in this process, causing the child to quickly become fatigued. This does not contribute to the effective development of the brain according to the age of the child, and the missed opportunities further hinder the development of the ability to perceive and process new information. The search for new ways of presenting knowledge that makes it possible to level negative trends, according to Feldstein (2013a), is one of the directions of the methodological and pedagogical scientific search for the near future. One of such methods is the use of the patterns of the unconscious component of the psyche in the process of perceiving educational information (Chernigovskaya, 2015; Chernigovskaya et al., 2016; Damasio, 2018; Dean, 2018; Filippova, 2006; Klementovich et al., 2016; Klochkov et al., 2008; Kryukova et al., 2017; Morozov & Spiridonov, 2019; Pervushina & Osetrin, 2017; Schenk, 2012; Verbitskaya, 2019).

2.2. The choice of a method for analyzing the transmission of meanings embedded in an array of educational information

For compatibility of educational information, taking into account its semantics, a system-generating contradiction is expressed in contrasting its logical coherence and isolation. Certain fields of knowledge are naturally associated with others, but their fragmentation in the process of cognition leads to some fragmentation of textbooks among themselves in the branches of the corresponding sciences. This determines the objective need to analyze the logical connectivity and sequence of educational information

on the one hand, and isolation, on the other (Klochkov et al., 2019; Rybakova et al., 2015; Rybakova et al., 2017).

The main property of educational information considered in the framework of this work is the semantic compatibility of its individual blocks in the sequence of perception provided for by the curriculum. The composition in this case can be considered as a set of iconic associations of various levels, presented in the form of a finished text; structure - as relations between them (Klochkov et al., 2019; Novikov, 2019; Rybakova et al., 2015; Rybakova et al., 2017; Tuchkova, 2019).

To determine the amount of information at the level of its semantic content (semantic level), the thesaurus measure (Groot et al., 2016; Lagutina et al., 2016; LeCun et al., 2015; Mai et al., 2017; Mai et al., 2018; Wilson et al., 2019). This characteristic determines semantic properties through the student's ability to accept (perceive and assimilate) the information received (Chernigovskaya et al., 2016; Kiselev, 2018; Popova, 2012). The concept of a thesaurus measure includes the concept of a thesaurus, which implies the totality of information available to the student or system. As applied to educational information, a thesaurus can be understood as a terminological component, which is expressed, as a rule, in the form of words or their combinations, concentrating the content in itself, which means that information compatibility can be determined up to the level of sentence analysis (Klochkov et al., 2019; Rybakova et al., 2015; Rybakova et al., 2017).

The educational information functions in the communication system between the knowledge carrier (*subject*) and its listener, those for whom it is intended as the ultimate addressee (subject). Information itself (*object*) is the link between the two entities and therefore carries the signs corresponding to its understanding by different parties. The knowledge carrier, laying a certain meaning (content) in the information, transforms its form in accordance with pedagogical goals, the main of which is to convey meaningful content with minimal distortion in the mind of the recipient (Klochkov et al., 2019; Rybakova et al., 2015; Rybakova et al., 2017).

2.3. Development of software for didactic analysis of semantic compatibility of texts

The main task of the software being developed will be to obtain an expanded thesaurus and to analyze publications for compatibility of information in them. This task requires the selection of a number of parameters of the algorithms, therefore, it assumes the presence of a developed interface that allows for research.

In modern conditions, services placed on the Internet are becoming more and more relevant, so the web-model of the application was chosen. Python 3.7 was chosen as the service language for the language model and text analysis. The interface was implemented using the Django (2020) framework.

3. Research Questions

During the development of the educational text processing program, the following questions were raised:

- choice of language model;
- what stages of pre-processing are relevant for the problem being solved;

- selection of vectorization of documents of the case, allowing to solve the task;
 - development of an effective algorithm for extracting a thesaurus;
- selection of parameters for the algorithm for extracting the subject thesaurus.

4. Purpose of the Study

It is assumed that the solution of the research problems reflected in the questions forms the main goal of the study is the development of a mathematical model and software for determining the extended thesaurus and analysis of documents for compatibility of information in them.

5. Research Methods

To highlight the thesaurus, we need to implement two main algorithms: the Stemming algorithm and the TF-IDF algorithm.

5.1. Text model and vector weighting methods

As a rule, the subject of a document is well described by the composition of the dictionary used in this document, as well as by the frequencies of words, and not by the semantic links between them. Therefore, for the task of highlighting a thesaurus containing subject vocabulary, models are usually used that work with particular frequency characteristics (Andreeva & Ushakov, 2019; Borodaschenko et al., 2015; Golitsyna et al., 2016; Kiselev et al., 2018; Metcalfe, 2017; Shenhav, 2017; Tsatsaronis et al., 2009). Models of this kind are conventionally referred to as “bag of words” (Bondarchuk, 2015).

They are characterized by a description of each analyzed document by a high-dimensional vector containing the weight of the word in the document. For the purpose of highlighting the thesaurus, two methods were chosen for weighting words: the frequency occurrence of words and the statistical measure TF-IDF.

In frequency vectorization, weight is the number of times a word is used in a document normalized to the L2 norm (or Euclidean norm) to smooth out the influence of the length of the document

$$n_{w_i} = \frac{w_i}{\sqrt{\sum_j w_j^2}}$$

where: w_i is the frequency of the i -th word.

TF-IDF statistical measure is used to analyze the significance of a word in the context of information contained in a text document that is part of a text array (collection, set of training texts) (Borodaschenko et al., 2015; Golitsyna et al., 2016; Kiselev et al., 2018 ; Tsatsaronis et al., 2009).

$$TF(w, d) = \frac{WordCount(w, d)}{Length(d)},$$

$$IDF(w, c) = \frac{Size(c)}{DocCount(w, c)},$$

$$TF-IDF(w, d, c) = TF(w, d) \times IDF(w, c),$$

где: $WordCount(w,d)$ is the frequency of w word in d document; $Length(d)$ is the length of d document; $Size(c)$ is case size; $DocCount(w,c)$ is the number of documents containing w word.

Those words in TF-IDF that have a high frequency and at the same time low frequency of use in other documents of the collection (Borodaschenko et al., 2015; Golitsyna et al., 2016; Kiselev et al., 2018; Tsatsaronis et al., 2009 and others) will have great weight, which allows you to adhere to a balance of frequency and information content.

5.2. Approaches to the formation of a subject thesaurus

The preliminary stage when using the “word bag” model is tokenization (allocation of tokens - words) and normalization (reduction of words to normal form), which contributes to a significant reduction in the dimension of vectors. In this work, normalization is performed using a lemmatizer, which is justified for the Russian language even for more complex models (Kutuzov & Kuzmenko, 2019).

Obviously, after the frequency vectorization and sorting of words by frequency, a distribution according to Zipf's law was obtained - a probability distribution describing the relationship between the frequency of an event and the number of events with such a frequency (Moreno-Sánchez et al., 2016; Piantadosi, 2014; Qiu et al., 2017; Yatsko, 2015; Zipf, 1936).

The most frequent words in the corpus are usually the least informative and rarely useful for any word processing tasks (Andreeva & Ushakov, 2019; Lagutina et al., 2019). In turn, low-frequency words are highly informative, however, they are unreliable as factors in decision-making. Thus, to highlight the subject thesaurus, it is necessary to take the middle part of the “rank-frequency” distribution.

Using the TF-IDF measure gives a different distribution, the most frequency words appear at the tail of the “TF-IDF” distribution. And to select the thesaurus, one need to select the words with the highest weights (Golitsyna et al., 2016; Kiselev et al., 2018; Tsatsaronis et al., 2009).

In the work, it is proposed to combine these two methods of obtaining the subject thesaurus, that is, to select words on the basis of a weighted average assessment of approaches, after which it is possible to analyze the increment of information and text compatibility.

The compatibility of training materials suggests that the growth of the thesaurus upon transition to a new topic does not exceed any threshold value (Golitsyna et al., 2016). Usually this value is chosen equal to 20%. That is, the growth of the thesaurus in the next chapter, the document should not exceed 15-20% (Metcalf, 2017; Shenhav et al., 2017; Wilson et al., 2019).

6. Findings

6.1. Software algorithm

In the course of work on the tasks set, the following general software operation algorithm was developed:

- pre-processing of the document: tokenization (word break); Lemmatization (performed using the mystem parser from Yandex); removal of stop words that do not carry semantic load

- vectorization and weighting using the normalized word frequency, obtaining the distribution of “rank - frequency” and the formation of the thesaurus as the middle part of the obtained distribution (threshold values are set by the researcher);
- vectorization and weighting of TF-IDF. obtaining the distribution of “rank - TF-IDF” and the formation of a thesaurus as the left side of the resulting distribution (the threshold value is set by the researcher);
- an expanded thesaurus is formed as a weighted average of the two received thesauri, an expert assessment is carried out;
- an analysis of the subject compatibility with other documents of the corpus is performed.

6.2. Software description

The software is developed on the basis of the Django framework, the mathematical model is implemented in python 3.7 using the scikit-learn library (2020).

The program interface allows loading a corpus of documents (chapters of one textbook or a set of textbooks on related disciplines).

After that, for each document of the corpus, the subject thesaurus is automatically selected with the specified selection parameters (selection boundaries from the frequency and TF-IDF distributions) and a list of words is formed - the subject thesaurus. At this stage, it is possible to review the thesaurus obtained and perform an expert assessment of the selection.

A separate module allows performing compatibility analysis of any two documents of the corpus, obtain distribution schedules, and also upload the results to a csv file for further work and analysis.

7. Conclusion

The software product was developed to analyze the compatibility of information in educational texts and is intended to be used as a tool for didactic analysis, both within the framework of a single document (text, file, textbook, manual), and their combination (set, collection, set), which allows adjust the flow of material and the sequence of studying sections, disciplines, taking into account the influence of the thesaurus increment speed.

Given the fact that a large proportion of educational texts for different levels of education (except for school) is compiled by specialists in their field who do not have sufficient knowledge from the point of view of the methodology and methodology of compiling textbooks, such a tool allows us to solve this problem. A smooth increment of the thesaurus within the required level of perception (Metcalfe, 2017; Shenhav et al., 2017; Wilson et al., 2019) will allow us to neutralize the negative trends in the change in the quality of perception of information (Feldstein, 2013a, 2013b), which is especially important when it is a digital transformation in the framework of continuing education (Verbitskaya, 2019).

Testing of the program for processing the collection of educational texts (on the example of educational materials on social studies and related disciplines) showed that the developed service copes with this task.

Due to the flexibility of the algorithms of the program, compatibility analysis can be performed on texts that are not educational, where analysis of their compatibility is required.

The development plans of the project include the development of additional modules to an existing system, for example, to automate the selection of model parameters, use distribution semantics methods for vectorizing texts, as well as optimizing the speed of algorithms.

References

- Andreeva, A. Yu., & Ushakov, B. K. (2019). Automation of the assessment of the validity of tests in electronic courses using natural language processing methods. *Measurement, control, informatization "IKI-2019"*, 146-150.
- Anokhin, K. V., & Velichkovsky B. M. (2011). Natural-science prospects of modern cognitive research *Vestnik of RFBR*, 2-3(70-71), 67-77.
- Bondarchuk, D. V. (2015). Algorithm for constructing a semantic core for the text classifier. *Siberian Journal of Life Sciences and Agriculture*, 8, 2.
- Borodaschenko, A. Yu., Potemkin, A. V., Sazonova, E. A., & Shekshuev, S. V. (2015). Algorithm for searching for similar media publications. *Science of Science*, 7(4), 64.
- Chernigovskaya, T. V. (2015). An experimental study of language and thinking in the XXI century: traditions and opportunities. *Proceedings of the Joint Scientific Council on humanitarian problems and historical and cultural heritage*, 64-76.
- Chernigovskaya, T. V. (2010). Reading in the context of cognitive knowledge. *On the ways to a new school*, 1, 11-13.
- Chernigovskaya, T. V., Shelepin, E. Yu., Zashirinskaya, O. V., & Nikolaeva E. I. (2016). *Psychophysiological and neurolinguistic aspects of the process of recognition of verbal and non-verbal communication patterns*. VVM.
- Damasio, A. (2018). *So begins "I". The brain and the emergence of consciousness*. Career Press.
- Dean, S. (2018). *Consciousness and the brain. How the brain encodes thoughts*. Career Press.
- Django (2020). *Django documentation*. <https://docs.djangoproject.com/en/3.0/>
- Feldstein, D. I. (2013a). Problems of forming the personality of a growing person at a new historical stage in the development of society. *Education and science*, 9(108), 3-22.
- Feldstein, D. I. (2013b). Problems of psychological and pedagogical sciences in the spatio-temporal situation of the XXI century (report at the general meeting of the Russian Academy of Education). *Russian Psychological Journal*, 10(2), 7-31.
- Filippova, M. G. (2006). *Investigation of unconscious perception (based on the material of multi-valued images)*. *Experimental psychology of cognition: cognitive logic of the conscious and unconscious*. Publishing House of St. Petersburg University.
- Golitsyna, O. L., Maksimov, N. V., & Fedorova, V. A. (2016). On determining semantic similarity based on relationships of a combined thesaurus. *Autom. Doc. Math. Linguist*, 50, 139–153. <https://doi.org/10.3103/S0005105516040026>
- Groot, F., Huettig, F., & Olivers, C. (2016). When Meaning Matters: The Temporal Dynamics of Semantic Influences on Visual Attention. *Journal of Experimental Psychology: Human Perception and Performance*, 42(2), 180–196.
- Kiselev, Y. A., Mukhin, M. Y., & Porshnev, S. V. (2018). *Automated Methods for Detecting Semantic Relations for Electronic Thesauri*. Goryachaya Liniya: Telekom.
- Klementovich, I. P., Levanova, E. A., & Stepanov, V. G. (2016) Neuropedagogy: a new branch of scientific knowledge. *Pedagogy and the psychology of education*, 2, 8-17.
- Klochkov V. P., Kamoza T. L., Krotova I. V., & Donchenko N. A. (2008). Analysis of the possibilities of the phenomenon of the unconscious in the educational process. *Humanitarian vector*, 1, 41-45.
- Klochkov, V. P., Barakhsanova, E. A., Krotova, I. V., Rybakova, G. R., & Malkova, T. V. (2019). Dichotomical indicators of educational information compatibility modeling. *Dilemas contemporáneos: Educación, Política y Valores*, 6(S8), 11.
- Kryukova, A. P., Agafonov, A. Yu., Kozlov, D. D., & Shilov, Yu. E. (2017). The effect of the asymmetry of semantic activation with an unconscious understanding of the multi-valued vocabulary. *Vestnik of Kemerovo State University*, 2, 158-163.
- Kutuzov, A., & Kuzmenko, E. (2019). To Lemmatize or Not to Lemmatize: How Word Normalisation Affects ELMo Performance in Word Sense Disambiguation. In *Proceedings of the First NLP*

- Workshop on Deep Learning for Natural Language Processing* (pp. 22-28). <https://www.aclweb.org/anthology/W19-6203>
- Lagutina, N. S., Lagutina, K. V., Mamedov, E. I., & Paramonov, I. V., (2016). Methodological aspects of semantic relation extraction for automatic thesaurus generation. *Model. Anal. Inf. Sist*, 23(6), 826–840.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436-444.
- Mai, F., Galke, L., & Scherp, A. (2018). Using Deep Learning for Title-Based Semantic Subject Indexing to Reach Competitive Performance to Full-Text. In *JCDL '18: Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries* (pp. 169–178). <https://doi.org/10.1145/3197026.3197039>
- Mai, F., Galke, L., Brunsch, D., & Scherp, A. (2017). Using Titles vs. Full-text as Source for Automated Semantic Document Annotation. In *K-CAP 2017: Proceedings of the Knowledge Capture Conference December, 20*, (pp. 1–4). <https://doi.org/10.1145/3148011.3148039>
- Metcalfe, J. (2017). Learning from errors. *Annu. Rev. Psychol*, 68, 465–489.
- Moreno-Sánchez, I., Font-Clos, F., & Corral, Á. (2016) Large-Scale Analysis of Zipf's Law in English Texts. *PLoS ONE*, 11(1), e0147073. <https://doi.org/10.1371/journal.pone.0147073>
- Morozov, M. I., & Spiridonov, V. F. (2019). Mechanisms for the influence of categorical information on visual search. *Vestnik St. Petersburg University. Psychology*, 9(3), 280–294.
- Morozova, L. V., & Novikova, Yu. V. (2013). Features of reading text from paper and electronic media. *Arctic Environmental Research*, 1, 81-86.
- Novikov A. I. (2019). Thesaurus as a reflection of the semantic space of a language. *Semantics of information technology*. http://it-claim.ru/Library/Books/Semantics_IT/before.htm
- Pervushina, N. A., & Osetrin, K. E. (2017). Neuropedagogy as an expression of the symbolism of bioethics *Scientific and pedagogical review*, 2(16), 198-208.
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions *Psychonomic bulletin & review*, 21(5), 1112-1130.
- Popova O. V. (2012). *Psycholinguistic modeling of the processes of perception of a scientific and educational text*. Kemerovo State University.
- Qiu, J., Zhao, R., Yang, S., & Dong, K. (2017). *Word Frequency Distribution of Literature Information: Zipf's Law Informetrics*. Springer. https://doi.org/10.1007/978-981-10-4032-0_5
- Rybakova G. R., Krotova I. V., & Kamoza T. L. (2015). Modeling the compatibility of educational information: methodological approaches *Humanitarian vector*, 1(41), 24-34.
- Rybakova, G. R., Andreeva, A. Yu., & Katkov, A. S. (2017). Program for modeling the compatibility of educational information "Thesaurus". *Certificate of registration of a computer program RU 2017662610*.
- Schenk, R. (2012). What should be the basis of learning - cognitive processes or substantive content? *Cognitive research*, 5, 289-290.
- Shenhav, A. (2017). Toward a rational and mechanistic account of mental effort. *Annu. Rev. Neurosci*, 40, 99–124.
- Tsatsaronis, G., Varlamis, I., Vazirgiannis, M., & Nørkvåg, K. (2009). Omiotis: A Thesaurus-Based Measure of Text Relatedness. In Buntine W., Grobelnik M., Mladenić D., Shawe-Taylor J. (Eds.), *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2009. Lecture Notes in Computer Science*, 5782. Springer.
- Tuchkova, N. P. (2019). Role and capabilities of specialized thesauruses in cognitive technologies *Information and mathematical technologies in science and management*, 1(13), 5-15. <https://doi.org/10.25729/2413-0133-2019-1-01>
- Verbitskaya, N. O. (2019). Digital transformation of continuing education: a new round in the development of neuropedagogy. *Bulletin of the South Ural State University*, 11(3), 6–20.
- Wilson, R. C., Shenhav, A., & Straccia, M. (2019). The Eighty Five Percent Rule for optimal learning. *Nat Commun*, 10, 4646. <https://doi.org/10.1038/s41467-019-12552-4>
- Yatsko, V. A. (2015). Automatic text classification method based on Zipf's law *Automatic Documentation and Mathematical Linguistics*, 49, 83-88.
- Zipf, G. K. (1936). *The Psycho-Biology of Language*. Routledge.