

SCTMG 2021
**International Scientific Conference «Social and Cultural Transformations in the Context of
Modern Globalism»**

**PYTHON LANGUAGE IN THE DEVELOPMENT OF
COMPUTATIONAL COMPETENCIES IN HUMANITIES
EDUCATION**

Anatoly B. Mantusov (a)*, Zamira B. Dorzinova (b)

*Corresponding author

(a) Kalmyk State University named after B.B. Gorodovikov, 11, Pushkina str., Elista, 358000, Russia,
mantab@yandex.ru,

(b) Kalmyk State University named after B.B. Gorodovikov, 11, Pushkina str., Elista, 358000, Russia,
zamira.mol@mail.ru

Abstract

This article examines the development of computational competencies in undergraduate and graduate students of humanities, studying philology through the use of free and open-source software such as the Python programming language on the example of quantitative text analysis using Python programming language libraries such as NumPy and SciPy, as well as Pandas. The leading platform for building Python programs to handle natural language data is NaturalLanguageToolkit, which is free and open-source software. For morphological analysis: obtaining the lemma or base of a given token and, if necessary, morphological parameters, we used the freeware program MyStem. Pandas is a package, an add-on to the NumPy library, that provides an efficient implementation of the DataFrame class. In Pandas we used structures like Series and Dataframe, and used the library matplotlib to visualize the result of cluster analysis as a dendrogram. The choice of training content concerning the application of these methods to arrays of textual data depends on the principles of professional orientation of training and ethno-cultural connotation of education, so the example of cluster analysis uses the Kalmykian folk epic "Dzhangar" and the sample questions to be proposed during the analysis of the example, choosing the object of subsequent application of the proposed method in the context of forming a professional picture of the world. In addition, a specific issue is the critical analysis of the processing of text analysis results obtained by applying the studied methods to further develop the teaching of computer science.

2357-1330 © 2021 Published by European Publisher.

Keywords: Python, computing competencies, ethno-cultural connotation, Dzhangar, Kalmykian folk epic



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

In our globalized and interconnected world, an essential factor determining the learning processes of the domestic vocational education system is developing competencies of future professionals, in particular competencies related to working with the information presented in different forms. Students of philology whose specialization involves studying culture through text will require an appropriate competency. In our opinion, textual competency is closely connected with computational competency, as it implies, in particular, computing with the text presented in its quantitative analysis, which on the one hand expands the concept of computational competency to the case of textual work, and on the other hand, the large amount of computing required when working with text determines the need for ICT. This implies an appropriate information competency, the choice of appropriate methods and tools, e.g. the type of software (free or proprietary) and programming language (C, Python, R).

2. Problem Statement

The problem of using software in shaping the competencies of future professionals has been relevant since the advent and widespread use of computers because dynamic changes in all areas of human activity and, in particular, in the computer sphere, constantly pose new challenges. In this paper we will limit our discussion to the use of FOSS such as the Python programming language (PL) in developing computational competencies for humanities education on the example of quantitative text analysis using Python libraries such as NumPy and SciPy.

3. Research Questions

This study examines the use of free and open-source software (FOSS) such as the Python programming language (PL) in developing computational competencies for humanities education on the example of quantitative text analysis using Python libraries such as NumPy and SciPy. We have attempted to answer the following questions:

- How should we organize the learning process to obtain computational competencies as a component of a future philologist's professional picture?
- How to achieve an effective learning experience through the use of free and open-source software?

Today all spheres of human activity, including education and philology, are using free software.

Various researchers have devoted their work to the use of free and open-source software in their professional work. Thus, Balandina et al. (2019) consider the problems that arise when using free software in the activities of teachers and suggests ways to overcome them. Kalyuzhny (2014) sees FOSS as a systemic factor in the information environment of science and society. Vanina et al. (2018) points out that for specialist training, it is best to use different combinations of FOSS products, with an expanded range of free software used and the use of universal systems that can run on any platform. Mantusov and Dorzhinova (2018, 2019a, 2019b) provide examples of the use of FOSS such as the LibreOffice Office Package and GraphViz. This study uses such research principles as a systematic approach to the analysis

of pedagogical phenomena; axiological and cultural approaches to the study of social and socio-pedagogical phenomena (M.N. Berulava, S.G. Vershlovsky, A.P. Valitskaya, V.P. Zinchenko, Y.N. Kulutkin, Y.M. Lotman, A.A. Orlov et al) and competency-based phenomenon (V.I. Baidenko, I.A. Zimnaya, E.F. Zeer, W. Hutmacher, R.W. White et al.).

The theoretical basis for the study was the work of the theory of higher education methodology. This time is characterized by recognizing the fact that professional knowledge, competencies, professional worldview are of paramount importance on the path to success. As a result, there is a significant increase in the importance of the principle of professional orientation, while the internationalization of education and increased academic mobility require increased attention to the principle of the ethno-cultural connotation of learning (Mantusov & Dorzhinova, 2019a, 2019b) The principle of professional orientation in higher education was introduced by Nizamov (1975) and Barabanshchikov (1976). The issues of vocational education and training have been researched in the works of Verbitsky (1991), Kudryavtsev (1981), Kuzmina and Tikhomirov (1972), Makhmutov (1985), Slastenin (1976) and others. Pankin (2002) introduced the principle of the ethno-cultural connotation of learning. The works of Tsurumova (2008), Vinogradova (2014) and Semenova (2005) have been investigated in the studies of the ethno-cultural orientation of education. Applying the principle of professional orientation in education and the principle of the ethno-cultural connotation in education is the methodological basis for optimizing the teaching and learning process in higher education institutions (Mantusov, 2007).

4. Purpose of the Study

The scientific and theoretical rationale provided the basis for the following objectives:

1. Provide students with a description of the use of Python in text analysis.
2. Provide students with examples of ways to use Python in text analysis.
3. Analyse whether the results of the cluster analysis of textual information can be interpreted.

5. Research Methods

To address these issues, we used a set of research methods: an analytical-synthetic review of scientific publications and theoretical works; method of analogy, modelling. In addition to the above-mentioned methods, we applied elements of content analysis, comparison, analogy and analysis of the results of educational activities.

6. Findings

Since one feature of the current period is the need for cross-platform availability, one solution to this problem is the use of the Python programming language which is distributed under the free Python SoftwareFoundationLicense (PSFL), a BSD-like permissive free software licence. This licence is compatible with the GNU GeneralPublicLicense (GPL). The PSFL licence defines the possibility to distribute the software project and this licence is non-copyleft and allows changes to the source code as

well as the creation of derivative works without opening the code. The PSFL allows the use of this language without restrictions in any application, including proprietary ones (Zozulya & Nechaev, 2018). Students were introduced to the theoretical premises and examined in practice various applications of the Python language and its packages, in particular SciPy for textual analysis. The leading platform for building Python programs to handle natural language data is NaturalLanguageToolkit (NLTK), NLTK is free software, the download address is <http://www.nltk.org>. We used Learning NLTK eBook (Learning NLTK eBook (PDF). Electronic resource // <https://riptutorial.com/ebook/nltk> date of access 10/09/2019) as an example. LearningnlkeBook is an unofficial and free NLTK eBook created for educational purposes contains examples of NLTK use in classification, tokenization, stemming, tagging, syntactic analysis and semantic parsing. The analysis involved the approach described in the work (Mantusov, 2019) and a software tool based on this approach to determine the frequency vocabulary of the text, where it is more convenient to use the freeware program MyStem, which performs the morphological analysis of text in Russian to obtain the lemma or base (pseudo-base) of a given token and, if necessary, morphological parameters, to perform morphological analysis. The download address is <https://yandex.ru/dev/mystem/>. Operation of the MyStem software is governed by the framework of the software licence agreement [MyStem software licence agreement / <https://yandex.ru/legal/mystem/>]. We used MyStem as an external library for Python "pymystem3", providing all the functionality of the programme to avoid connecting external applications. To create frequency dictionaries, we used the Counter class of the Python standard library, from the collections library. Then, after creating frequency dictionaries of the texts under study, we use SciPy, an open-source library for the Python programming language (official website: <https://www.scipy.org>), (official documentation is available at <https://docs.scipy.org/doc/scipy/reference/tutorial/io.html>) designed for scientific and engineering calculations, in our case, for cluster analysis. To ensure that data analysis and modelling tasks work within Python, we have used pandas, a Python software library for data processing and analysis which is distributed (official website: <https://pandas.pydata.org>). Pandas provides an efficient implementation of structures such as Series and DataFrame, among others. Using Pandas allows the typical steps of data processing and analysis — loading, preparation, manipulation, modelling and analysis. We used Pandas structures like Series and Dataframe. Series is a structure to work with a sequence of univariate data, in our case it is used for the results of frequency analysis of each of the proposed tests, and Dataframe is used for structuring the results presented as a Series in the process of cluster analysis. We used the matplotlib library to visualize the results of cluster analysis as a dendrogram. Matplotlib is distributed under a BSD-like licence, a short guide to Matplotlib is available at https://pyprog.pro/mpl/mpl_short_guide.html. Guided by the principle of professional orientation, the students have to consider both the methods used and the results of the calculations and their interpretation, for example: describe changes in the resulting dendrogram when changing such parameters as the stemming, the function used to calculate the distance (Distance Metric to be used. The distance function can be 'braycurtis', 'canberra', 'chebyshev', 'cityblock', 'correlation', 'cosine', 'dice', 'euclidean', 'hamming', 'jaccard', 'jensenshannon', 'kulsinski', 'mahalanobis', 'matching', 'minkowski', 'rogerstanimoto', 'russellrao', 'seuclidean', 'sokalmichener', 'sokalsneath', 'sqeuclidean' etc). Applying the described methods as implementing the principle of professional orientation of learning and the principle of the ethno-

The study of the dendrogram raises questions about the reasons for such a grouping, into a distinct group of songs within the third song, the twelfth song and the thirteenth song, which in itself could be the subject of a separate study and comparative analysis of different versions of the Kalmyk folk epic Dzhangar. It is possible to raise similar questions using the Kazakh epic "Koblandy-batyr", the Altai epic "Maadai-Kara", the Yakut epic "Olonkho", the Karelian-Finnish epic "Kalevala" and the Lezgi epic "Sharvili" as objects of study, Bashkir epic "Ural-batyr", Evenki epic "Nimngakan" and many other literary monuments depending on the specific situation as the implementation of the principles of the ethno-cultural connotation of education and professional orientation of training.

7. Conclusion

The need for computational competency becomes a necessary skill that opens up access to successful studies and is crucial to professional success. These competencies will open up a world of information, prospects, opportunities for young people to navigate through life and live in a dynamic information society.

This study aims to find out ways of organizing the development of computational competencies in humanities students.

Humanities students with computational competencies have the skills:

- (1) name and apply the theoretical foundations for analysing different texts, especially relating to different literary monuments,
- (2) explain and critically evaluate their country's culture,
- (3) identify, classify and critically compare different information presented in text form,
- (4) identify and communicate their personal experience of applying computing competencies,
- (5) classify texts based on their cluster analysis.

Students are invited to take part in various tasks. Their involvement contributes to an understanding of which elements of IT applications for frequency analysis of texts are particularly useful. The study notes the great methodological impact of the use of information technology in text analysis. The findings show an increase in computational competency and will contribute to the design of Python-based applications under development for a wider range of tasks.

Acknowledgments

We thank our families for their support.

References

- Balandina, I. V., Osokina, E. V., & Maslikova, T. A. (2019). Problems in the use of free software in teachers' activities and ways of overcoming them. *A World of Science. Pedagogy and Psychology*, 5. <https://mir-nauki.com/PDF/77PDMN519.pdf>
- Barabanshchikov, V. (1976). Some theoretical and practical issues of higher school pedagogy. In R. A. Nizamov (Ed.), *Modern Problems of Higher School Pedagogy* (pp. 4–30). Publishing house of Kazan University.

- Kalyuzhny, K. A. (2014). Free software as a system-forming factor of information environment of science and society: state and prospects. *Science. Innovation. Education*, 16. <https://cyberleninka.ru/article/n/svobodnoeprogrammnoe-obespechenie-kak-sistemoobrazuyuschiy-faktor-informatsionnoysredy-nauki-i-obschestva-sostoyanie-i-perspektivy>
- Kudryavtsev, A. Ya. (1981). To the problem of learning principles. *Soviet pedagogy*, 8, 100–106.
- Kuzmina, N. V., & Tikhomirov, S. A. (1972). Methodological problems of university pedagogics. In *Problems of Higher School Pedagogy* (pp. 6–43). Leningrad.
- Makhmutov, M. I. (1985). The principle of professional orientation of training. In *Principles of learning in modern pedagogical theory and practice*. ChPU.
- Mantusov, A. B. (2007). *Educational technologies in optimizing the teaching and learning process in higher education*. Kalmyk State University named after B.B. Gorodovikov.
- Mantusov, A. B. (2019). *The formation of a scientific picture of the world of future economists in the process of studying the educational field of "Mathematics"*. Kalmyk State University named after B.B. Gorodovikov".
- Mantusov, A. B., & Dorzhinova, Z. B. (2018). Formation of computational competencies in future economists while using LIBREOFFICE CALC programme. *Vector of Science of Togliatti State University. Series: Pedagogy, Psychology*, 4(35), 26–33.
- Mantusov, A. B., & Dorzhinova, Z. B. (2019a). One approach to automatic quantitative text analysis using the Python language on the example of the Kalmyk epos "Dzhangar". In *Network Orientalism: interaction of Mongolian and Turkic ethnic groups in time and space. Proceedings of the III International Scientific Forum* (pp. 265–267). Elista.
- Mantusov, A. B., & Dorzhinova, Z. B. (2019b). Using visualization with GRAPHVIZ software in forming computational competencies. *Business. Education. Right*, 3(48), 384–392.
- Nizamov, R. A. (1975). *Didactic bases of activation of students' learning activity*. KSU.
- Pankin, A. B. (2002). *Designing national-regional educational systems based on the principle of the ethno-cultural connotation* [Doctoral Dissertation]. Kalmyk State University.
- Semenova, A. D. (2005). Directions for improving ethno-pedagogy of the educational process in a rural national school. *Bulletin of Yakutsk State University*, 2(2), 105–110.
- Slastenin, V. A. (1976). *The formation of the personality of a Soviet school teacher in the process of professional training*. Prosveshchenie.
- Tsuryumova, S. V. (2008). Ethno-cultural connotations of supplementary art and aesthetic education for children. In A.V. Zolotareva (Ed.), *Extended Education for Children as a Factor in the Development of Educational Institutions of Various Types: Proceedings of the All-Russian Scientific-Practical Correspondence Conference* (pp. 137–140). YaSPU Press.
- Vanina, M. F., Davydova, E. V., Erohin, A. G., & Frolova, E. A. (2018). Problems and Prospects of Russian and Foreign Free Software Use in the Educational Process of Higher Education. *Methodological issues of infocommunication teaching in higher education*, 7(1), 7–11.
- Verbitsky, A. A. (1991). *Active learning in higher education: a contextual approach*. Vysshaya Shkola.
- Vinogradova, N. V. (2014). Ethno-cultural connotation in conditions of additional education for children. *Education and upbringing: methods and practice*, 12, 94–100.
- Zozulya, M. M., & Nechaev, S. A. (2018). Free software on Python language platform as a part of informational and educational environment. *NovaUm.Ru*, 15, 420–423.