20 Septembre, 2024;
Paris, République française

SECTION 14.
PHILOLOGIE ET JOURNALISME

DOI 10.36074/logos-20.09.2024.031

BIG DATA IN PHILOLOGY

Svitlana Krasniuk[1], Svitlana Goncharenko[2]

1. Senior Lecturer, Department of philology and translation
Institute of Law and Modern Technologies Kyiv National University of Technologies and Design, UKRAINE
ORCID ID: 0000-0002-5987-8681

2. Senior Lecturer
Institute of Law and Modern Technologies Kyiv National University of Technologies and Design, UKRAINE
ORCID ID: 0000-0002-7740-4658

Introduction.

Data in philology significantly changed the traditional methods of researching languages, texts and literature, thanks to the use of computer technologies and methods of processing large volumes of textual information. Within digital philology, scientists actively use digital tools and methods to analyze texts, study language changes, and study cultural and social aspects of languages.

Big Data are data arrays that, in terms of their volume, variety and speed of generation, exceed the capabilities of traditional methods of information collection, storage, processing and analysis. Big data has become a central component of the modern digital era, affecting a wide range of fields, such as science, education, business, medicine, education, government, and others [1-5].

But by themselves, accumulated and stored Big Data do not produce additional and new usefulness, do not generate added value, new knowledge and new insights. It is the analysis and analytics of big data that are the central components of the modern process of working with large volumes of information [5, 6]. They allow organizations and researchers to identify hidden regularities and patterns, formalize new knowledge, make predictions and make informed decisions [7, 8].

Machine learning (ML) and Big Data are closely related and form the foundation of modern innovations in many fields, including philology. Big data provides a large amount of information for training machine learning models, while machine learning provides efficient methods for processing and analyzing this

data, allowing to obtain valuable information and predictions [9-12].

Deep Learning and Big Data are two key technologies that mutually reinforce each other, creating new opportunities for information processing and analysis [13, 14]. Their integration provides significant achievements in various fields of science and industry, from medicine to finance, from technology to social sciences and humanities, including linguistics.

Big data in philology has a significant impact on modern research in the field of linguistics, literary studies, textology, and other disciplines. The use of large arrays of text, audio, and video data provides philologists with new opportunities for the analysis of linguistic and literary phenomena, allowing the study of significant volumes of information that were previously inaccessible to traditional research methods [15, 16]. This opens up prospects for new scientific discoveries in languages, literature and culture.

Big data in linguistics.

Linguistics is one of the key areas of philology, where the use of big data allows for detailed and multidimensional studies of languages. Large volumes of text corpora available for analysis help researchers to carry out the following tasks:

1. Corpus linguistics.

Corpus linguistics uses large arrays of texts (corpora) to study patterns in linguistic data. Important aspects of research in this field are:

- Frequency analysis: Ability to determine the frequency of use of words and grammatical structures in large text databases. This allows conducting research not only at the level of individual works, but also at the level of the language system as a whole.

- Collocational analysis: Study of combinations of words and phrases to understand the contextual use of language elements. The use of big data makes it possible to identify collocations that are rare but significant in certain contexts.

- Language evolution: Big data allows tracking changes in language over time. This provides an opportunity to explore historical changes in vocabulary, grammatical rules, and stylistic trends.

2. Dialectology and sociolinguistics.

Big data helps in the study of language variants and dialects by analyzing data from social networks, blogs, forums and other sources. This provides an opportunity to obtain massive samples for the analysis of speech characteristics of different social, age and geographical groups. For example, language differences between different regions or social groups can be identified based on vocabulary and grammar analysis.

3. Lexicography.

Large text databases became the basis for modern lexicography, helping to

create dictionaries based on real language data. The analysis of large corpora makes it possible to discover new words, changes in the meanings of existing words, and the emergence of new meanings.

Big data in literary studies.

The use of big data in literary studies makes it possible to conduct quantitative analyzes of texts and study literary phenomena at a new level. This applies both to the analysis of individual authors and works, and to broad studies of literary movements and trends.

1. Digital analysis of literature.

Digital humanities actively uses big data to analyze literary texts. The main approaches are:

- Stylistic analysis: By analyzing large volumes of texts, researchers can study the peculiarities of the style of individual authors or literary schools. It helps to determine the characteristic features of the author's style, differences between genres or changes in the work of one author over time.

- Themes and motifs: Big data allows you to automatically classify texts by themes and motifs, which facilitates the study of literary trends in different eras or regions. For example, one can trace how motifs such as love, war, or death have changed in different literary traditions.

2. Sentiment analysis of literature.

Big data makes it possible to perform automated analysis of the emotional tone of literary works, which allows researchers to study emotional patterns in the works of different authors or genres. This becomes especially useful when studying historical changes in literature or comparing authors within the same period.

3. Network analysis of literary characters.

One of the interesting areas is the analysis of relationships between characters in works of literature. Using big data, it is possible to build networks of interaction of characters in novels, which allows analyzing social structures and dynamics of character development in literary works. This approach allows for quantitative comparisons between different works and authors.

Big Data in Textology and Authorship Analysis.

Textology is a branch of philology that deals with the study of texts, their variants, and establishing the authenticity of texts. Big data can be used for automated analysis of text variants and authorship detection. This includes:

- Attribution of authorship: Using statistical methods to determine the authorship of a text based on stylistic and lexical characteristics. For example, by analyzing the frequency of use of certain words or grammatical structures, it is possible to identify the authors of anonymous or dubious works.

- Textual analysis: Big data allows analyzing the variability of texts, for example,

different versions of literary works or manuscripts. This helps textologists trace the evolution of texts and identify original redactions.

Challenges and prospects of using big data in philology.

Despite significant opportunities, the use of big data in philology is accompanied by certain challenges:

- Technical limitations: Big data requires significant computing resources for storage and processing. Most philological research requires special tools and platforms for analyzing texts.

- Data quality: Not all texts are available in a structured and clean format. This especially applies to handwritten or ancient texts that may be damaged or presented in different versions.

- Data interpretation: Big data enables quantitative metrics, but its interpretation still requires deep knowledge of context, culture, and history.

Conclusions.

The integration of data into philological research has significantly changed approaches to the study of languages and texts. The use of large amounts of textual and linguistic information allows for deeper and more accurate analyzes of language phenomena, to reveal new regularities and to make scientific discoveries in the field of linguistics, literary studies, and cultural studies.

Big data has become an important component of the modern information society, providing new opportunities for analysis and decision-making. Their use has the potential to significantly influence the development of various industries, but at the same time requires solving numerous technical, ethical and legal issues.

Big data analysis and analytics enable organizations to extract valuable information from vast amounts of data, make predictions and make strategic decisions. The development of data processing technologies, such as machine learning and cloud computing, is constantly improving the capabilities of analysis, which opens new horizons for the application of big data in various fields of activity.

Machine learning and big data mutually reinforce each other, creating opportunities for automated analysis of large volumes of information, uncovering hidden patterns, and improving decision-making. These technologies have enormous potential in various industries, from medicine to finance, and continue to evolve with technological advances.

The integration of deep machine learning with big data provides new opportunities for analyzing and processing information, which opens the horizons for significant achievements in many areas. However, in order to realize the potential of these technologies, it is necessary to solve a number of technical, resource and ethical problems. The further development of data processing technologies and the improvement of machine learning architectures promise

even more improvements in the processing and use of big data.

Big data in philology provides new opportunities for research, allowing quantitative analysis of linguistic and literary phenomena on a large-scale level. They become an indispensable tool for corpus analysis, lexicography, attribution of authorship and digital analysis of literature. At the same time, the use of big data opens up new methodological challenges that require the development of tools and approaches for their effective use.

In the future, philological research will increasingly rely on collaboration between philologists and information technology specialists. The rise of artificial intelligence, in particular deep learning, opens up new opportunities for automatic translation, handwriting recognition, as well as for the analysis of rare languages. The development of open text data platforms will facilitate research by making access to historical and cultural texts faster and more convenient.

Discussion and prospects for further research.

The authors emphasize the significant prospects of applying hybrid machine learning to big data in the field of linguistics and philology. Hybrid machine learning in linguistics combines various machine learning approaches and techniques to achieve better results in various complex and complex tasks [17-20]. This approach combines traditional machine learning methods with modern deep learning techniques, leveraging their advantages to improve the accuracy, adaptability and efficiency of the models. It is hybrid machine learning in linguistics that takes advantage of different approaches to achieve more accurate and efficient results. The combination of traditional methods with modern deep models allows you to create powerful systems for natural language processing, which opens up new opportunities for automating and improving the analysis of textual information. The use of such hybrid approaches provides greater flexibility, accuracy and adaptability in various linguistic tasks.

This direction of scientific research of the authors will be reflected in the following publications.

REFERENCES:

[1] Науменко, М. (2024). Аналіз та аналітика великих даних в маркетингу та торгівлі конкурентного підприємства. Grail of Science, (40), 117–128. https://doi.org/10.36074/grail-of-science.07.06.2024.013.
[2] Maxim Krasnyuk, Svitlana Nevmerzhytska, Tetiana Tsalko. (2024). Processing, analysis & analytics of big data for the innovative management. Grail of Science, #38, April 2024. pp. 75-83. https://www.journal-grail.science/issue38.pdf.
[3] Maxim Krasnyuk, Dmytro Elishys (2024). Perspectives and problems of big data analysis & analytics for effective marketing of tourism industry. Science and technology today, #4 (32) 2024. pp. 833-857.

[4] Krasnyuk M., Krasnuik I. Big data analysis and analytics for marketing and retail. Штучний інтелект у науці та освіті: збірник тез Міжнародної наукової конференції (AISE) (1-2.03.2024 р.), Київ, 2024.
[5] Krasnyuk M.T., Hrashchenko I.S., Kustarovskiy O.D., Krasniuk S.O. (2018) Methodology of effective application of Big Data and Data Mining technologies as an important anti-crisis component of the complex policy of logistic business optimization. Economies' Horizons. 2018. No. 3(6). pp. 121–136.
[6] Kulynych Y., Krasnyuk M., Krasniuk S. Knowledge discovery and data mining of structured and unstructured business data: problems and prospects of implementation and adaptation in crisis conditions. Grail of Science. 2022. (12-13). pp. 63-70.
[7] Краснюк Світлана. (2024) Data Science у освітньому менеджменті // Діалог культур у Європейському освітньому просторі [Електронний ресурс]: Матеріали IV Міжнародної конференції, м. Київ, 10 травня 2024р. Київський національний університет технологій та дизайну / упор. С. Є. Дворянчикова. – К.: КНУТД, 2024. – С. 119-124.
[8] Tetiana Tsalko, Svitlana Nevmerzhytska, Svitlana Krasniuk, Svitlana Goncharenko, Liubymova Natalia «Features, problems and prospects of data mining and data science application in educational management» // Bulletin of Science and Education, №5(23), 2024. pp.637-657.
[9] Krasnyuk M., Krasniuk S. Comparative characteristics of machine learning for predicative financial modelling. ΛΌГОΣ. 2020. P. 55-57.
[10] Krasnyuk M., Tkalenko A., Krasniuk S. Results of analysis of machine learning practice for training effective model of bankruptcy forecasting in emerging markets. ΛΌГОΣ. 2021.
[11] Науменко, М. (2024). Ефективне застосування класичних алгоритмів машинного навчання при прийнятті адаптивних управлінських рішень. Наукові перспективи, 2024, #5 (47). https://doi.org/10.52058/2708-7530-2024-5(47)-855-875.
[12] Krasnyuk M., Krasniuk S. Modern practice of machine learning in the aviation transport industry. ΛΌГОΣ. 2021.
[13] Науменко, М. (2024). Оптимальне використання алгоритмів глибокого машинного навчання в ефективному управлінні підприємством. Успіхи і досягнення у науці, 2024, #4 (4). https://doi.org/10.52058/3041-1254-2024-4(4)-776-794.
[14] Maxim Krasnyuk, Svitlana Krasniuk, Svitlana Goncharenko, Liudmyla Roienko, Vitalina Denysenko, Liubymova Natalia. Features, problems and prospects of the application of deep machine learning in linguistics // Bulletin of Science and Education, №11(17), 2023. pp.19-34. http://perspectives.pp.ua/index.php/vno/article/view/7746/7791.
[15] Krasniuk, S., & Goncharenko, S. (2024). Ethics of using large language models in machine linguistics. In Лінгвістичні та методологічні аспекти викладання іноземних мов професійного спрямування. Національний авіаційний університет.
[16] Goncharenko, S., & Krasniuk, S. (2024). Innovative architecture of large language models. In Лінгвістичні та методологічні аспекти викладання іноземних мов професійного спрямування. Національний авіаційний університет.
[17] Краснюк М.Т. Гібридизація інтелектуальних методів аналізу бізнесових даних (режим виявлення аномалій) як складовий інструмент корпоративного аудиту. Стан і перспективи розвитку обліково-інформаційної системи в Україні: матеріали III Міжнар. наук.-практ. конф. (м. Тернопіль, 10-11 жовт. 2014 р.). Тернопіль: ТНЕУ, 2014. С. 211-212.
[18] Гращенко І.С., Краснюк М.Т., Краснюк С.О. Гібридно-сценарне застосування

інтелектуальних, орієнтованих на знання технологій, як важливий антикризовий інструмент логістичних компаній в Україні. Вчені записки Таврійського Національного Університету імені В. І. Вернадського. Серія: Економіка і управління. 2019. Т. 30 (69). С. 121-129.

[19] Krasnyuk M., Goncharenko S., Krasniuk S. Intelligent technologies in hybrid corporate DSS. Інноваційно-інвестиційний механізм забезпечення конкурентоспроможності країни: колективна монографія / за заг. ред. О. Л. Гальцової. Львів-Торунь: Ліга-Прес, 2022. С. 194-211.

[20] Krasnyuk M., Hrashchenko I., Goncharenko S., Krasniuk S. Hybrid application of decision trees, fuzzy logic and production rules for supporting investment decision making (on the example of an oil and gas producing company). Access to science, business, innovation in digital economy. ACCESS Press. 2022. 3(3). P. 278-291.