ABSTRACT

Explanatory note size – 143 pages, contains 33 illustrations, 32 tables, 8 applications, 65 references.

Topicality. Examines the problem of the problem of processing a large amount of text data in the context of plagiarism detection using Big Data solutions. The main features of modern platforms for identifying loans are analyzed, their advantages and disadvantages are determined. The need to improve natural language processing methods applied to text data to increase the efficiency and accuracy of plagiarism detection has been revealed.

The aim of the study. The main target is increasing the speed and reliability of plagiarism detection software on extremely large data sets.

The object of research: plagiarism detection software.

The subject of research: methods, software architecture, tools for creating software for detecting plagiarism on large amounts of data.

To achieve this goal, the **following tasks** were formulated:

- analysis of existing solutions;
- selection and preparation of a dictionary of Ukrainian and English synonyms;
- creation of an updated method of detecting plagiarism using dictionaries of synonyms;
- software architecture creation and implementation of software using big data solutions;
- update of the morphological library for the Ukrainian language;
- assessment of the effectiveness of the proposed solution.

The scientific novelty of the results of the master's dissertation is in proposed an improved plagiarism check method using a thesaurus and added support for scalability and handling of extremely large datasets. The result was achieved by developing a modernized software architecture.

The practical value of the obtained results is in developed a modified method and software that can be used in academic institutions and conferences to ensure academic integrity and reduce the use of borrowed texts.

Relationship with working with scientific programs, plans, topics. Work was performed at the Department of Informatics and Software Engineering of the National Technical University of Ukraine «Kyiv Polytechnic Institute. Igor Sikorsky».

Approbation. The scientific provisions of the dissertation were tested at the all-Ukrainian competition of student works 2023, Kyiv.

Publications. The scientific provisions of the dissertation were published in:

1) Halaiko D.O., Oliynyk Y.O. SEARCHING TEXT SIMILARITY PARALLEL METHOD. SoftTech-2022, 2022 #004.414.
2) Yurii Oliinyk, Danylo Halaiko, Iryna Mukha, Oleksandr Ocheretianyi Plagiarism detecting hash-based parallel method. Proceedings of the 7th international conference, COLINS-2023. 2023. Vol. IV. P. 131–143.
3) Halaiko D.O., Oliynyk Yu.O. Application of data warehouses to detect plagiarism in text documents. Adaptive systems of automatic management, 2024, Volume 2 #45.

Keywords: BORROWING, PLAGIARISM, TEXT, BIG DATA, BIG DATA, DYNAMODB, AWS, FIREBASE.