ABSTRACT

Explanatory note size – 94 pages, contains 17 illustrations, 25 tables, 5 applications, 29 references.

Topicality. Examines the problem in the field of text data stream processing, shows the main features of existing text data stream processing platforms, their advantages and disadvantages. The need to improve the natural language processing method for text data streams has been identified.

The aim of the study. The main goal is to improve existing natural language processing tools to provide support for the Ukrainian language and to develop software capable of analyzing text data streams in real time.

The object of research: methods of text data stream processing.

The subject of research: methods and means of creating software for natural language processing for text data streams in real time.

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

- comparative analysis of available solutions for processing text data streams in real time;
- formulation of structural and technical features of sources of text data flows;
- selection and preparation of a Ukrainian dictionary;
- implementation of existing solutions to support morphological analysis;
- development of software for processing text data streams using a morphological analyzer in real time;
- assessment of the effectiveness of the proposed solution.

The scientific novelty of the results of the master's dissertation is is that the improvement of the natural language processing method of text data due to the introduction of real-time stream processing support, which increases the speed of

processing and allows to perform distributed calculations, as well as the improvement of the level of support for the Ukrainian language due to the integration of the WESUM dictionary.

The practical value of the obtained results is that the software architecture for real-time processing of text data streams using Apache Spark and the Spark Streaming library with storage of results in the Elasticsearch search server using the Kibana visualization engine is proposed, and software is developed using of the proposed architecture for the analysis of streams of Ukrainian-language textual data. The developed software can be used in the future to process streams of text data from Ukrainian-language sources, as well as to perform a wider range of NLP tasks (for example, sentiment analysis or intellectual analysis of text data).

Relationship with working with scientific programs, plans, topics. The work was performed at the Department of Informatics and Software Engineering of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" in the framework of the topic "Methods and technologies of high-performance computing and big data processing". State registration number 0117U000924.

Approbation. The scientific provisions of the dissertation were tested at the Third All-Ukrainian Scientific and Practical Conference of Young Scientists and Students "Software Engineering and Advanced Information Technologies" (SoftTech-2022 autumn) - Kyiv.

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

1) Fedorovych I.A. Processing models of text data streams in the Apache Spark Structured Streaming engine / I.A. Fedorovych, Yu.O. Oliinyk // Materials of the III All-Ukrainian scientific and practical conference of young scientists and students "Software engineering and advanced information technologies" (SoftTech-2022 autumn) - Kyiv: NTUU "KPI them. Igor Sikorsky", November 23-25, 2022.

Keywords: NATURAL LANGUAGE PROCESSING, TEXT DATA STREAM PROCESSING, REAL-TIME STREAM PROCESSING, APACHE SPARK, SPARK STRUCTURED STREAMING.