

ABSTRACT

Master's dissertation consists 104 pages, 39 images, 26 tables, 3 appendix, 21 referring sources.

Topicality: The adaptation of NLP methods to Ukrainian texts remains an unresolved challenge. Therefore, the creation of a high-quality and unified method for analyzing mixed direct speech (Ukrainian and Russian, as well as Surzhyk and slang expressions) is urgently needed.

The purpose of the dissertation research is to adapt methods for analyzing direct multilingual speech for application in Ukrainian-language chats, enhance the quality of analyzing direct speech containing Russianisms and slang expressions.

Object of study: Approaches and processes of creating software for classifying and categorizing social media and messenger messages.

Subject of research: Methods, models, software architecture, tools for creating software for analysis, classification, and categorization of multilingual and slang-enriched direct speech.

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

- analyze existing solutions for analyzing direct multilingual speech;
- analyze and adapt existing methodologies for processing multilingual text for application to Slavic languages, specifically Ukrainian and Russian;
- improve and adapt methods for analyzing multilingual data for application to Slavic languages, particularly Ukrainian and Russian, considering the processing of slang expressions in the modern Ukrainian language;
- develop architecture and software with implemented methods for analyzing multilingual data;
- develop a plan for scaling and deploying the software in the cloud;
- research the effectiveness of the developed methods and software.

Scientific novelty: within the scope of the work, existing methods for processing Ukrainian direct speech were improved, and methods for classifying direct speech containing Russianisms, Surzhyk, and slang expressions were formalized and adapted.

The practical value of the obtained results is determined by the fact that developed software library will enable the automation of the classification and categorization of multilingual messages from social networks and messengers. It also guarantees high efficiency when applied to texts in multilingual Slavic languages, specifically Ukrainian and Russian, as well as to surzhyk and slang expressions.

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 conference "Informatics and Computer Engineering - SoftTech-2023".

Publications: Theses of the thesis are published in «Informatics and Computer Engineering SoftTech-2023».

Keywords: NLP, TOPIC MODELING, CLASSIFICATION, BAG OF WORDS MODEL, NEURAL NETWORKS.