

ABSTRACT

Explanatory note size – 87 pages, contains 32 illustrations, 24 tables, 5 applications, 78 references.

Topicality. Examines the problem of interaction with data sources using natural language instructions, shows the main features of existing solutions for the described problem, their advantages and disadvantages. A need has been identified for software to prepare datasets from multiple sources using natural language instructions.

The aim of the study. The main target is to improve the process of interaction with databases and data repositories in the way of developing software that implements a convenient interface to databases and data repositories using natural language commands.

The object of research: interfaces to databases and data storages using natural language.

The subject of research: models, methods and technologies for processing commands in natural language in order to build an interface for preparing sets of data from several sources.

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

- analysis of the subject area, research, publications and actual problems of developing interfaces based on natural language;
- analysis of existing software solutions, their advantages and disadvantages;
- formation of requirements for the software product;
- architecture design and technology selection for software product development;
- software product development;
- software product testing and analysis of the obtained results.

The scientific novelty of the results of the master's dissertation is that the use of the capabilities of LLM models for the preparation of data sets from several sources with natural language commands has been further developed. The result is achieved by using Large Language Model models for processing instructions in natural language, and a data

warehouse approach for combining data from several sources into a common SQL interface.

The practical value of the obtained results is the development of a web application capable of preparing datasets from multiple sources using natural language instructions. This application combines data sources into a common SQL interface using the data warehouse pattern. The application is able to process voice and text commands in English, convert them into SQL to the corresponding unified interface, execute the generated query and provide the execution results to the user in CSV format.

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» within the framework of the initiative topic "Methods and technologies of high-performance computing and processing of ultra-large data sets", registration number UKRINTEI 0117U000924

Approbation. The scientific provisions of the dissertation were tested at the scientific conferences SoftTech 2022 and SoftTech 2023.

Keywords: DATA STORAGE. NATURAL LANGUAGE PROCESSING, DATA WAREHOUSE, NATURAL LANGUAGE INTERFACE.