АНОТАЦІЯ

Бакалаврський дипломний проєкт включає пояснювальну записку (53 стор., 30 рис., список використаної літератури з 23 найменувань, 3 додатка).

Об'єктом розробки даного проєкту є програмний продукт для автоматизованого аналізу Big Data та виявлення дублікатів. Його призначення полягає в забезпеченні ефективної та точної обробки великого обсягу даних з метою виявлення дублікатів за різними критеріями.

У процесі розробки програмного продукту було проведено наступні кроки:

- Аналіз вимог до програмного продукту і формулювання вимог до нього.
- Розробка алгоритмів обробки даних та виявлення дублікатів.
- Розробка та тестування програмного продукту.
- Внесення необхідних змін та вдосконалення програмного продукту на основі результатів тестування.

Результатом розробки програмного продукту є забезпечення швидкого та точного аналізу великого обсягу даних з метою виявлення дублікатів. Областю подальшого використання можуть бути компанії та організації, що працюють з великим обсягом даних та потребують їх ефективного аналізу.

Ключові слова:

BIG DATA, АВТОМАТИЗОВАНИЙ АНАЛІЗ ДАНИХ, ВИЯВЛЕННЯ ДУБЛІКАТІВ, АЛГОРИТМИ, ЕФЕКТИВНІСТЬ, ТОЧНІСТЬ.

ABSTRACT

The Bachelor's thesis project includes an explanatory note (53 pages, 30 figures, a list of references with 23 titles, 3 appendices).

The object of this project is a software product for automated Big Data analysis and duplicate detection. Its purpose is to provide effective and accurate processing of a large volume of data to identify duplicates based on various criteria.

The following steps were taken in the process of developing the software product:

- Analysis of requirements for the software product and formulation of requirements for it.
- Development of data processing algorithms and duplicate detection.
- Development and testing of the software product.
- Making necessary changes and improvements to the software product based on test results.

The result of the software product development is providing fast and accurate analysis of a large volume of data for duplicate detection. Companies and organizations that work with a large volume of data and require their efficient analysis may use it.

Keywords:

BIG DATA, AUTOMATED DATA ANALYSIS, DUPLICATE DETECTION, ALGORITHMS, EFFICIENCY, ACCURACY.