

АНОТАЦІЯ

Магістерська дисертація присвячена актуальному питанню автоматизації перевірки структурованих документів на плагіат. Головною метою роботи є пошук ефективного алгоритму для перевірки структурованих документів на плагіат.

В магістерській дисертації ретельно досліджені алгоритми для перевірки текстів на плагіат, формалізовано та реалізовано процес перевірки структурованих документів на плагіат. За основу аналізу тексту в документі взятий алгоритм «Шинглів». Алгоритм був модифікований реалізацією відокремлення з документу зображень та пошуку їх дублікатів в локальній базі користувача.

В роботі було розкрито теоретико-методичні та програмні основи, які були використані під час дослідження та знаходження ефективного рішення. Таке рішення ефективно вплинуло на кінцевий результат, тобто на кінцевий відсоток виявлення запозичень.

Ключові слова: плагіат, алгоритм, шингл, супершингл, мегашингл, структурований документ, канонізація.

Розмір пояснювальної записки – 87 аркушів, містить 36 ілюстрацій, 6 додатків.

ABSTRACT

The master's dissertation is devoted to the actual issue of automation of the verification of structured documents for plagiarism. The main goal of the work is to find an effective algorithm for verifying structured documents for plagiarism.

In the master's thesis carefully studied algorithms for checking texts on plagiarism, formalization and implementation of the process of verifying structured documents for plagiarism. The basis of the analysis of the text in the document is the algorithm "Shingles". The algorithm was modified by the implementation of separating from the document of images and finding their duplicates in the local user base.

The theoretical, methodological, and programmatic foundations that were used during research and finding an effective solution were revealed in the paper. Such a decision effectively influenced the final result, that is, the final percentage of borrowing detection.

Keywords: plagiarism, algorithm, shingle, supershingle, megashingle, document structure, canonization.

The size of the expository note is 87 sheets. It consists of 36 illustrations, 6 appendices.