

РЕФЕРАТ

Актуальність теми. У епоху інформації людство змушено зберігати велику кількість інформації через постійне її накопичення, що є великою проблемою через обмеженість пам'яті у наших девайсах. Наразі кількість файлів може бути не великою, але роздільна здатність зображень, відео зумовлює дороговартісне зберігання, передачу інформації через мережі. Найбільш популярним є алгоритм Гаффмана, що використовує техніку без втрат, тобто відбувається повне відновлення файлу до його першопочаткового вигляду після декодування. Стиснення даних також використовується для зберігання даних при резервному копіюванні, архівуванні, видаленні надлишковості.

Об'єктом дослідження є вивчення принципів роботи з даними та ознайомлення з жадібним алгоритмом Гаффмана.

Предметом дослідження оптимізація алгоритму Гаффмана у пам'яті, що в результаті призводять до його пришвидшення роботи.

Мета роботи: покращити використання пам'яті та часу роботи алгоритму за рахунок побудови оптимального Н-дерева, що конструюється на вхідних ймовірностях зустрічі певних символів у вхідному файлі.

Наукова новизна: запропонований спосіб стиснення даних, що створений на базі алгоритму Гаффмана та модифікований за допомогою застосування лишнього біту через особливості людської мови та реалізації її у носіях, що збільшує швидкість обходження Н-дерева.

Практична цінність: алгоритм навіть сьогодні застосовується в стисненні відео, аудіо та звичайних зображень, а в даній роботі, цінність результатів полягає в тому, що запропоновані методи дають змогу зменшити використання пам'яті при передаванні даних, як наслідок пришвидшити цей процес. При оцінці результатів ми отримали, що дерево Гаффмана для

англійського тексту взятого з книги зменшилося близько від 40 до 52 відсотків в залежності від вхідних даних.

Апробація роботи. Основні положення та результати роботи представлені та обговорювались на XV науковій конференції магістрантів та аспірантів «Прикладна математика та комп'ютинг» ПМК-2022 (Київ, 16-18 листопада 2022 р.). Опублікована стаття на конференції «IX Міжнародна науково-технічна Internet-конференція «Сучасні методи, інформаційне, програмне та технічне забезпечення систем керування організаційно-технічними та технологічними комплексами» (Київ, 25 листопада 2022 р.).

Структура та обсяг роботи. Магістерська дисертація складається з вступу, чотирьох розділів та висновків.

У вступі подано загальну інформацію щодо проблеми сучасного світу, обґрунтовано актуальність дослідження, сформульовано поставлені цілі та задачі, показано наукову новизну та практичну цінність роботи, наведено відомості про апробацію результатів і їхнє впровадження.

У першому розділі розглянута теорія існуючих алгоритмів стиснення даних, проведений їх аналіз відносно вибраного методу.

У другому розділі проведений аналіз оцінок методів стиснення, описано алгоритм дослідження, визначені основні переваги та недоліки, наведена теоретична частина модифікації алгоритму Гаффмана.

У третьому розділі наведена архітектура програми побудованої для дослідження результатів вдосконалення алгоритму Гаффмана на мові C++/Qt, описані технології, які були використані під час створення програми.

У четвертому розділі проведений аналіз результатів застосування методу оптимізації пам'яті в наявному алгоритмі з оригінальним алгоритмом та іншими найбільш розповсюдженими алгоритмами.

У висновках представлені результати проведеної роботи.

Робота представлена на 80 аркушах, наведено 5 таблиць, прикріплено 44 рисунки та наявні посилання на список використаних літературних джерел.

Ключові слова: оптимізація дерева; Н-дерево; жадібний алгоритм Гаффмана; стиснення даних без втрат.

ABSTRACT

Actuality of theme. In the information age, humanity is forced to store a large amount of information due to its constant accumulation, which is a big problem due to the limited memory in our devices. At present, the number of files may not be large, but the resolution of images and videos requires expensive storage and transmission of information over networks. The most popular is the Huffman algorithm, which uses a lossless technique, the file is completely restored to its original form after decoding. Data compression is also used for data storage during backup, archiving, and deduplication.

The object of research is a study of the principles of working with data and familiarization to the greedy Huffman algorithm.

The subject of the study is an optimization of the Huffman algorithm in memory, which results in its acceleration.

The goal of the work: to improve the use of memory and the running time of the algorithm due to the construction of an optimal H-tree, which is constructed on the input probabilities of meeting certain symbols in the input file.

The scientific novelty: the proposed method of data compression, created on the basis of the Huffman algorithm and modified by using an extra bit due to the peculiarities of human language and its implementation in data storage, which increases the speed of traversing the H-tree.

Practical value: the algorithm is still used today in the compression of video, audio and ordinary images, and in this work, the value of the results lies in the fact that the proposed methods make possible to reduce the use of memory during data transmission, as a result, to speed up this process. When evaluating the

results, we found that the Huffman tree for the English text taken from the book was reduced by about 40 to 52 percent, depending on the input data.

The scientific novelty the use of an extra bit due to the peculiarities of human language and its implementation in data storage, which increases the speed of traversing the H-tree, which is built on the probability table of the frequency of symbols in the input data set.

Approbation of work. The main provisions and results of the work were presented and discussed at the XV scientific conference of undergraduates and graduate students «Applied Mathematics and Computing» PMK-2022 (Kyiv, November 16-18, 2022). The article was published at the conference «IX International Scientific and Technical Internet Conference» Modern Methods, Information, Software and Technical Support of Management Systems of Organizational, Technical and Technological Complexes» (Kyiv, November 25, 2022).

Structure and scope of work. The master's dissertation consists of an introduction, four chapters and conclusions.

The introduction provides general information about the problem of the modern world, substantiates the relevance of the research, formulates the set goals and objectives, shows the scientific novelty and practical value of the work, provides information about the approbation of the results and their implementation.

The first chapter, the theory of existing data compression algorithms is considered, and their analysis is carried out in relation to the chosen method.

The second chapter, an analysis of compression method evaluations is carried out, the research algorithm is described, the main advantages and

disadvantages are determined, and the theoretical part of the modification of the Guffman algorithm is presented.

The third chapter, the architecture of the program built to study the results of improving the Guffman algorithm in the C++/Qt language is given, the technologies that were used during the creation of the program are described.

The fourth chapter analyzes the results of applying the memory optimization method in the existing algorithm with the original algorithm and other most common algorithms.

The results of the work are presented in the conclusions.

The work is presented on 80 sheets, 5 tables are given, 44 pictures are attached and there are links to the list of used literary sources.

Keywords: tree optimization; H-tree; Huffman's greedy algorithm; lossless data compression.