

УДК004.627

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ І СТИСКАННЯ ДАНИХ РІЗНИМИ АЛГОРИТМАМИ

Ковальов Костянтин Віталійович, Дікусар Катерина Володимирівна
к.т.н., доцент каф. КІСМ Шапорін Володимир Олегович
Національний університет “Одеська політехніка”, УКРАЇНА

АНОТАЦІЯ. Ця робота присвячена дослідженню ефективності стиснення даних різними алгоритмами. У роботі були використані алгоритми стиснення даних, такі як: LZW, Huffman, Arithmetic Coding та RLE. Для оцінки ефективності використання алгоритмів було обрано різні види даних, включаючи текстові дані, зображення та відео. Значення ефективності було визначено шляхом обчислення різними методами. Також було протестовано файли різних об’ємів (розміру).

Вступ. В даний час стискання даних є найважливішою складовою інформаційних технологій, які широко використовуються в різних галузях - від зберігання даних на жорстких дисках до прискорення передачі даних по мережі. Існує безліч алгоритмів стискання даних, кожен з яких має свої переваги та недоліки. У роботі проведено дослідження ефективності різних алгоритмів стискання даних на різних наборах даних. Результати цього дослідження допоможуть визначити, який алгоритм найкраще підходить для певних типів даних та завдань.

Для отримання результатів дослідження було обрано кілька популярних алгоритмів стискання, таких як LZ77, алгоритм Huffman та інші. Кожен алгоритм був застосований до набору вхідних даних, щоб оцінити його ефективність. Для кожного алгоритму буде виміряно ступінь стискання, час стискання та час розпакування даних.

Мета роботи-полягає у вивченні різних алгоритмів стискання даних [1] та їх ефективності на різних типах даних. У процесі роботи будуть розглянуті як класичні алгоритми, такі як LZW та Huffman, так і більш сучасні методи. Результати роботи дозволять проаналізувати ефективність різних алгоритмів на різних типах даних та визначити найбільш підходящі методи для конкретних завдань стискання даних.

Основна частина роботи. В рамках цього дослідження буде перевірено ефективність різних алгоритмів, а саме швидкість стискання, час стискання, наведено псевдокоди алгоритмів, наведено блок-схеми алгоритмів.

Попередньо будуть тестуватися наступні алгоритми:

- RLE;
- LZ77;
- LZ78;
- LZSS;
- LZW;
- LZAri;
- Huffman;
- Huffm anadaptive;
- Arithmetic adaptive;
- Інші алгоритми.

Дослідження ефективності стискання даних різними алгоритмами – це важливе завдання у галузі комп’ютерних наук та інформаційних технологій.

Існує безліч алгоритмів стискання даних (рис 1), кожен з яких має свої переваги та недоліки. Мета дослідження ефективності стискання даних – знайти оптимальний алгоритм стиснення для конкретного типу даних.

Для проведення дослідження можна використовувати різні метрики, такі як коефіцієнт стискання, час стискання та розпакування, ступінь стискання та якість зображення або звуку після стискання. Існують програми, що дозволяють порівнювати ефективність роботи різних алгоритмів на конкретних вхідних даних.

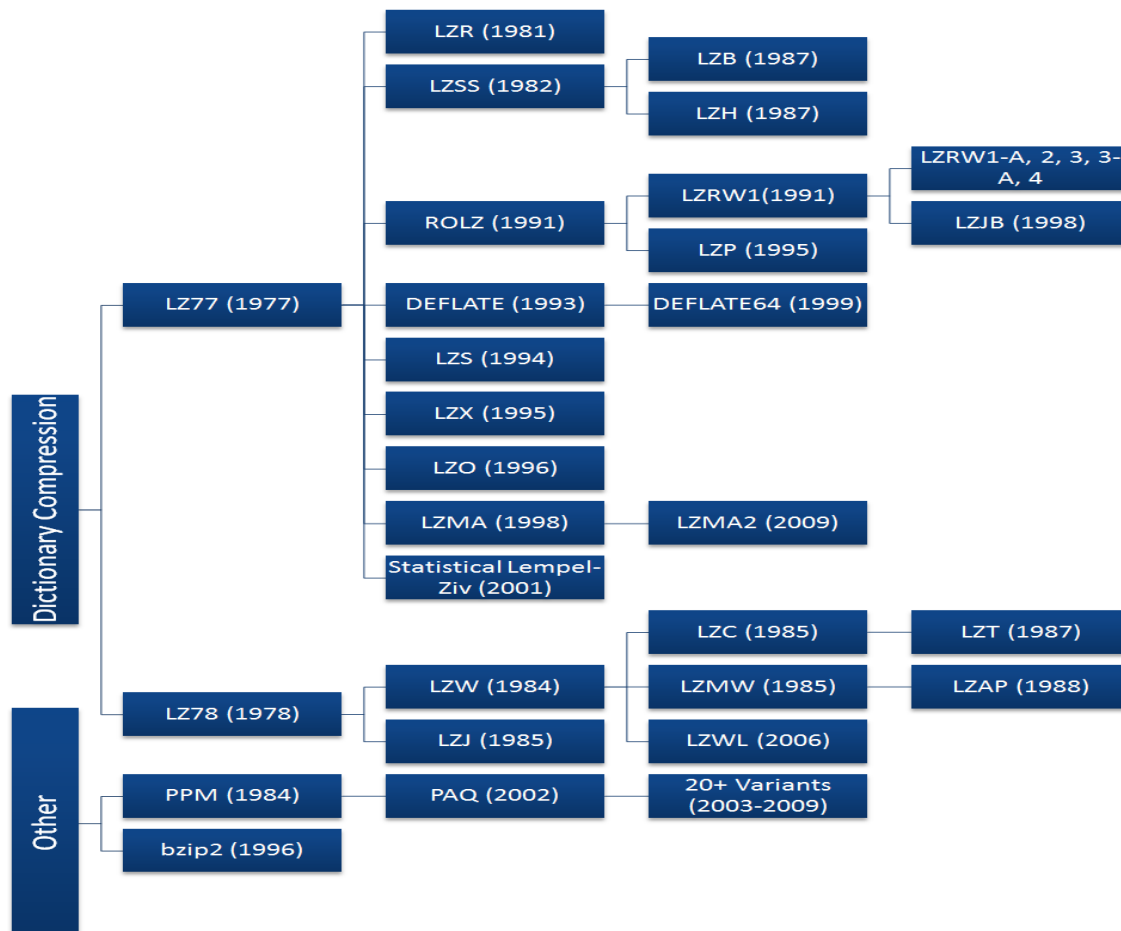


Рисунок 1 – Розвиток алгоритмів стискування 1977 -2011[2]

Дослідження ефективності стискування даних має безліч практичних застосувань, таких як оптимізація обсягу даних, що передаються в інтернеті або на пристроях зберігання даних, прискорення роботи програм, що працюють з великими обсягами даних і багато іншого.

Алгоритми можуть вимагати різної кількості ресурсів обчислювальної системи, на яких вони реалізовані. Чим ефективніший і універсальніший алгоритм, тим більші вимоги до обчислювальних ресурсів він висуває. Тим не менш, у спеціальних випадках прості та компактні алгоритми можуть працювати не гірше за складні та універсальні. Системні вимоги визначають їх споживчі якості: що менш вимогливий алгоритм, тим простіший, отже, компактною, надійною і дешевою системою може бути реалізований.

Один з них – це операція з стискування методом Huffman (на рис. 2 зображено його порівняння з оптимальним стискуванням), який є одним із найвідоміших і найпоширеніших алгоритмів стискування даних. Він був розроблений в 1951 Девідом Хаффманом і дозволяє досягти оптимального стиснення даних. Цей алгоритм полягає в тому, що кожному символу з вихідного файлу надається код, який залежить від частоти його появи в цьому файлі. Чим частіше з'являється символ, тим коротше код йому надається.

Також відомо про інші алгоритми стискування даних, наприклад, алгоритми Лемпеля-Зіва, які були розроблені в 1977 і 1984 роках відповідно. Ці алгоритми використовуються для стискування текстових даних і дозволяють досягти вищого ступеня стискування в порівнянні з алгоритмами Хаффмана. Вони засновані на пошуку повторюваних послідовностей символів у тексті та заміні їх на посилання на попередні входження цих послідовностей.

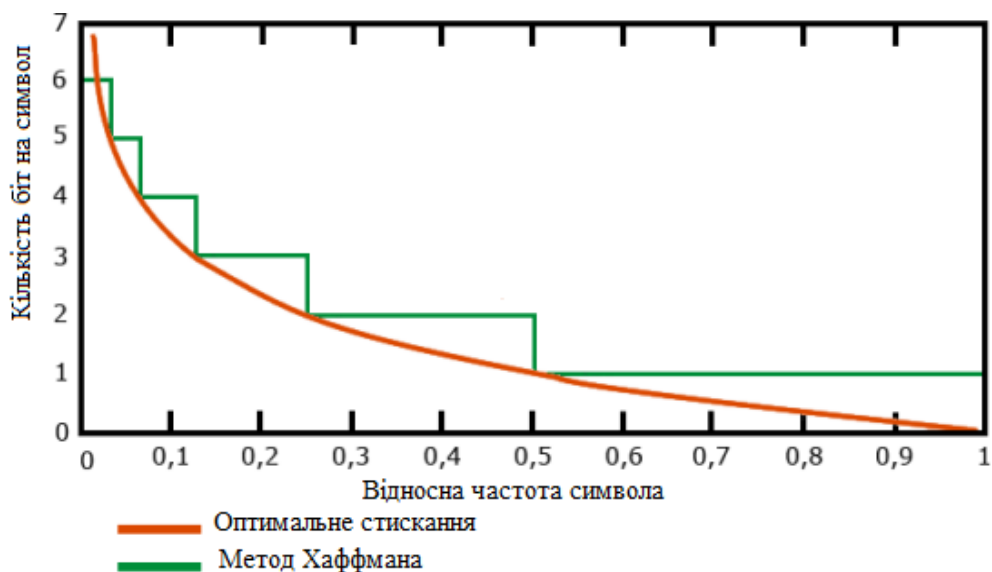


Рисунок 2 – Метод Хаффмана проти оптимального стиснення

Висновки. Дослідження ефективності стиснення даних різними алгоритмами є важливим завданням у галузі комп'ютерних наук та інженерії. Для стиснення даних можна використовувати різні алгоритми, такі як LZ77, LZSS та багато інших. При виборі алгоритму стиснення необхідно враховувати різні фактори, такі як тип даних, тип сховища, доступність апаратних ресурсів, швидкість стиснення/розпакування та багато інших. Ефективність різних алгоритмів може сильно змінюватись в залежності від цих факторів. Для дослідження ефективності стиснення даних можна використовувати різні метрики, такі як коефіцієнт стиснення, швидкість стиснення і розпакування, споживання пам'яті і т.д. Часто дослідження виконуються на наборах даних різного розміру та структури, щоб отримати повніше уявлення про продуктивність алгоритмів. Дослідження ефективності стиснення даних може бути корисним при виборі алгоритму стиснення для конкретної програми або додатків, тощо.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Wikipedia . -Електронний ресурс].-Режим доступу:
https://ru.wikipedia.org/wiki/%D0%A1%D0%B6%D0%B0%D1%82%D0%B8%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85.
2. History of Lossless Data Compression Algorithms-стаття.-[Електронний ресурс].
Режим доступу :https://ethw.org/History_of_Lossless_Data_Compression_Algorithms.

RESEARCH OF DATA COMPRESSION EFFICIENCY BY VARIOUS ALGORITHMS

Kostiantyn Kovalov, Kateryna Dikumar

PhD, Associate Professor of the department of CISND Shaporin Volodymyr
Odessa Polytechnic National University, UKRAINE

ABSTRACT. This paper is devoted to the research of the efficiency of data compression by various algorithms. Data compression algorithms such as LZW, Huffman, Arithmetic Coding, and RLE were used in this paper. Various types of data were selected to evaluate the effectiveness of the algorithms, including text data, images, and videos. The efficiency value was determined by calculation using various methods. Files of different sizes were also tested.