

ПОРІВННЯ РОЗРОБЛЕНОГО АЛГОРИТМУ НЕЛІНІЙНОГО СТИСНЕННЯ З ІСНУЮЧИМИ АЛГОРИТМАМИ НЕЛІНІЙ- НОГО СТИСНЕННЯ ДАНИХ

¹ Вінницький національний технічний університет;

Анотація

У цій роботі порівнюється шість методів стиснення, і на основі результатів доведено, що розроблений алгоритм може бути використаний як заміна загального алгоритму deflate.

Порівняно продуктивність розробленого алгоритму шляхом вимірювання ступеня стиснення та швидкості, а також швидкість декомпресії на трьох різних Canterbury corpus.

Ключові слова: стиснення, deflate, lzma, Canterbury corpus

Abstract

This paper compares six compression methods, and based on the results it is proved that the developed algorithm can be used as a replacement for the general deflate algorithm.

The performance of the developed algorithm was compared by measuring the compression ratio and velocity, as well as the decompression rate on three different Canterbury corpus.

Keywords: compression, deflate, lzma, Canterbury corpus.

Вступ

Значна частина практичного стиснення даних без втрат виконується не тільки за допомогою алгоритму deflate тому що він добре підтримується існуючими системами, а також тому, що він відносно простий і швидкий для кодування та декодування. МСА (розроблений алгоритм) - це власно побудований алгоритм стиснення і приклад реалізації цього алгоритму, який принципово ефективніший, ніж deflate. У цій роботі вимірюється ефективність власної реалізації та порівняння її з deflate та з кількома іншими алгоритмами стиснення.

Результати дослідження

Тести проводились із 22-бітним розміром вікна для МСА, LZMA та LZHAM та 15-бітовим вікном розмір для zopfli та d. Був використаний 22-бітний розмір вікна, оскільки минулі дослідження показали, що більші вікна можуть бути повільнішими для декодування. Більші розміри вікон, як правило, дають більший ступінь стиснення за рахунок швидкості декодування. Для deflate та zopfli було використано максимальний розмір вікна, дозволений форматом.

Версії перевірених алгоритмів: MCA version 0.1.0b, deflate algorithm from zlib 1.2.8 [1], Zopfli version from github 2015-09-0, LZMA implementation in 7zip 9.20.1 [3], and bzip2 1.0.6, 6-Sept-2010 [4].

Було використано такий тестовий стенд, процесор Intel Core i7-10750H, що працює на частоті 4 GHz, з шести ядрами в дванадцяти потоках. Усі алгоритми було скомпільовано за допомогою одного ж того самого компілятора GCC 4.8.4 при оптимізації рівня O2, всі тести було виконано в однопотоковому режимі.

Набір даних для стиснення, які використовувались при тестуванні, - Canterbury corpus [4] та enwik8, який використовується в премії Хаттера [5]. Середній розмір файлу в збірнику веб-вмісту становить лише 55 кБ, тому більша перевага вікна в розмірі нових алгоритмів над deflate в основному зникає там.

Було виміряно ступінь стиснення, швидкість стиснення та швидкість декомпресії для вибраних алгоритмів та рівнів стиснення. Швидкість стиснення і декомпресії кожного алгоритму виміряно за

допомогою тієї ж базової програми. Було обмежено вибір алгоритмів тими, які, як правило, мають вищий ступінь стиснення ніж у deflate. З цієї причини виключено алгоритми, такі як lz4 та zstd, з цього дослідження.

Таблиця 1 - Результати алгоритму стиснення Canterbury corpus (збірник включає 11 файлів)

Алгоритм: налаштування якості	Стиснення	Швидкість стиснення (МБ/сек)	Швидкість декомпресії (МБ/сек)
MCA:5	4.347	98.3	354.5
deflate:5	2.913	35.6	347.3
Zopfli	3.580	0.2	342.1
lzma:5	3.952	3.9	116.0
bzip2:5	3.869	12.0	40.4

Таблиця 2 - Результати різних алгоритмів стиснення у файлі enwik8

Алгоритм: налаштування якості	Стиснення	Швидкість стиснення (МБ/сек)	Швидкість декомпресії (МБ/сек)
MCA:5	3.308	78.3	279.4
deflate:5	2.742	18.1	217.4
zopfli	2.857	0.6	227.7
lzma:5	3.696	3.44	71.8
bzip2:5	3.447	12.4	30.3

Висновки

В результаті аналізу даних у таблиці 1 та у таблиці 2, можна стверджувати, що використання розробленого алгоритму є доцільним для невеликих файлів, що являється його основним призначенням, так як він розроблявся для використання на веб-серверах для збільшення швидкості отримання даних. А результати роботи цього алгоритму з великими файлами суттєво не відрізняються від інших.

Результати показують, що розроблений алгоритм нелінійного стиснення даних з усіх порівняних алгоритмів, буде вдалою заміною алгоритму deflate у трьох головних аспектах: коефіцієнт стиснення, швидкість стиснення та швидкість декомпресії.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. David, Salomon (2007). Data Compression: The Complete Reference (4 ed.). Springer. p. 241. ISBN 978-1-84628-602-5
2. Maydaniuk V. P. Increasing the Speed of Fractal Image Compression Using Two-Dimensional Approximating Transformations / V. P. Maydaniuk, I. R. Arseniuk, O. O. Lishchuk // Journal of Engineering Sciences. – Sumy : Sumy State University, 2019. – Volume 6, Issue 1. – P. E16 – E20.
3. Арсенюк І. Р. Комп'ютерні мережі: навчальний посібник / І. Р. Арсенюк, А. А. Яровий – Вінниця: ВНТУ, 2010 – 145 с.
4. Ian H. Witten; Alistair Moffat; Timothy C. Bell (1999). Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kaufmann. p. 92. ISBN 9781558605701.
5. Marcus Hutter, Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability, Springer, Berlin, 2004,

Горобець Юрій Володимирович — аспірант групи AC-19, факультет інформаційних технологій та комп'ютерної інженерії, кафедра комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, e-mail: yurii.sparrow@gmail.com

Перевозніков Сергій Іванович — доктор. техн. наук, професор кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, e-mail: perevoznikov@ukr.net

Озеранський Володимир Сергійович — канд. техн. наук, ст. викладач кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, e-mail: ozeransky@ukr.net

Horobets Yuri V. — graduate student of AC-19 group, Faculty of Information Technologies and Computer Engineering, Department of Computer Science, Vinnytsia National Technical University, Vinnytsia, e-mail: yurii.sparrow@gmail.com

Ozeranskyj Volodymyr S. - lecturer of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia, email: ozeransky@urk.net.

Perevoznikov Sergiy I. - professor of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia, email: perevoznikov@urk.net.

Науковий керівник: *Перевозніков Сергій Іванович* — доктор. техн. наук, професор кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, e-mail: perevoznikov@ukr.net