Презентація

Archiver Revolution

Кодування/Декодування

Перетворення Берроуза-Уіллера (*Burrows-Wheeler Transform*) з'явилося досить нещодавно (в *1994* році), ідеї використовувати сортування в алгоритмах компресії з'явилися ще в *80*-х, проте тоді методу не було знайдено застосування.

Сам по собі, *BWT* не є алгоритмом стиснення, однак вихідна послідовність набагато зручніше для стиснення, ніж вихідна.

Алгоритм діє блочно, професіонали в області стиснення рекомендують використовувати розмір буфера *1-2 KB*. Потім отриманий блок піддають циклічним перестановок, записуючи результати в список.

У результаті виходить наступне:

математика
атематикам <
тематикама <
ематикамат
матикамате
атикаматем <
тикаматема <
икаматемат
каматемати
аматематик

Зауважимо, що два слова, що починаються на "а", закінчуються на "м". Ще два слова – аналогічна пара: "т" – "а". Зблизимо ці рядки. Для цього застосуємо сортування в лексикографічному порядку.

Ми отримали такий список:

аматематик
атематикам
атикаматем
ематикамат
икаматемат
камателмати
математика <-- 6
матикамате
тематикама
тикаматема

Виберемо з цього списку вихідний рядок, а також збережемо всі заключні літери. У результаті ми отримали: 6кммттиаеаа.

Зворотне перетворення

Давайте запишемо дані, які в нас залишилися:

.....к

.....м

.....м

.....т

.....т

.....и

.....а

.....е

.....а

.....а

Так як відбувалося циклічні перетворення, то у нас залишилися букви в тій же кількості, в якій вони були в слові. Також ми знаємо, що оригінальна матриця була лексикографічно відсортована, значить, якщо ми відсортуємо літери, які у нас залишилися, то отримаємо вірну послідовність: кммттиаеаа -> аааеикммтт

а.....к
а.....м
а.....м
е.....т
и.....т
к.....и
м.....а
м.....е
т.....а
т.....а

Далі складаються послідовності, що складаються з останньої букви і початку рядка, сортуються, і вбирає відповідні літери
ка, ма, ма, те, ти, ик, ам, ем, ат, ат
ам, ат, ат, ем, ик, ка, ма, ма, те, ти

ам.....к
ат.....м
ат.....м
ем.....т
ик.....т
ка.....и
ма.....а
ма.....е
те.....а
ти.....а

І так далі. У кінцевому підсумку ми отримуємо оригінальну послідовність.

Стиснення і розтиснення

Кодування довжин серій (англ. *Run-length encoding, RLE*) або Кодування повторів - простий алгоритм стиснення даних, який оперує серіями даних, тобто послідовностями, у яких один і той же символ зустрічається кілька разів поспіль. При кодуванні рядок однакових символів, що складають серію, замінюється рядком, яка містить сам повторюється символ і кількість його повторів.

Як приклад наведена якийсь довільний рядок зображення в чорно-білому варіанті. Тут *B* представляє чорний піксель, а *W* позначає білий.:

WWWWWWWWWWWWWWWWBWWWWWWWWWWWWWWWWBBBWWWWWWWW
WWWWWWWWWWWWWWWWWWWWWWWWBWWWWWWWWWWWWWWWWWWW

Якщо ми застосуємо просте кодування довжин серій до цього рядку, то отримаємо наступне.:

12W1B12W3B24W1B14W

Таким чином, код представляє вихідні *67* символів у вигляді всього лише *18*.

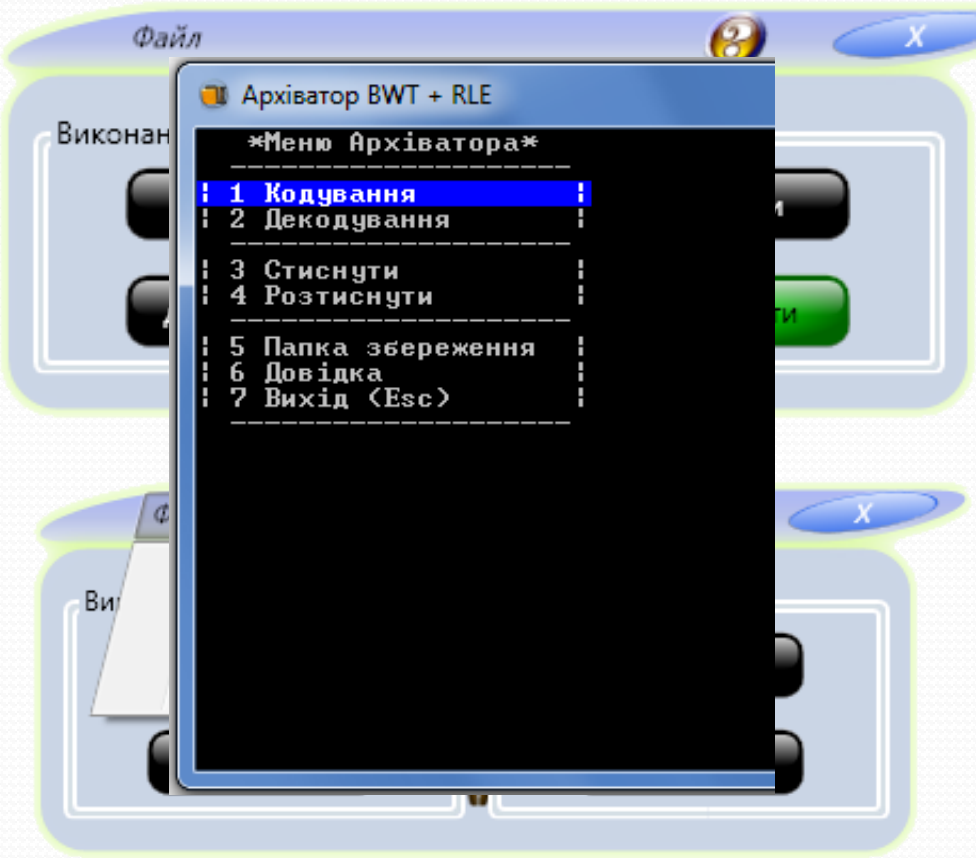
Однак, у разі, якщо рядок складається з великої кількості неповторюваних символів, її обсяг може зрости.

ABCABCABCABCDDEFFFFFFFFF

1A1B1C1A1B1C1A1B1C1A1B1C2D1E8F

Запис на деякій мові програмування алгоритму *RLE* з урахуванням цих обмежень нетривіальна.

Головний інтерфейс



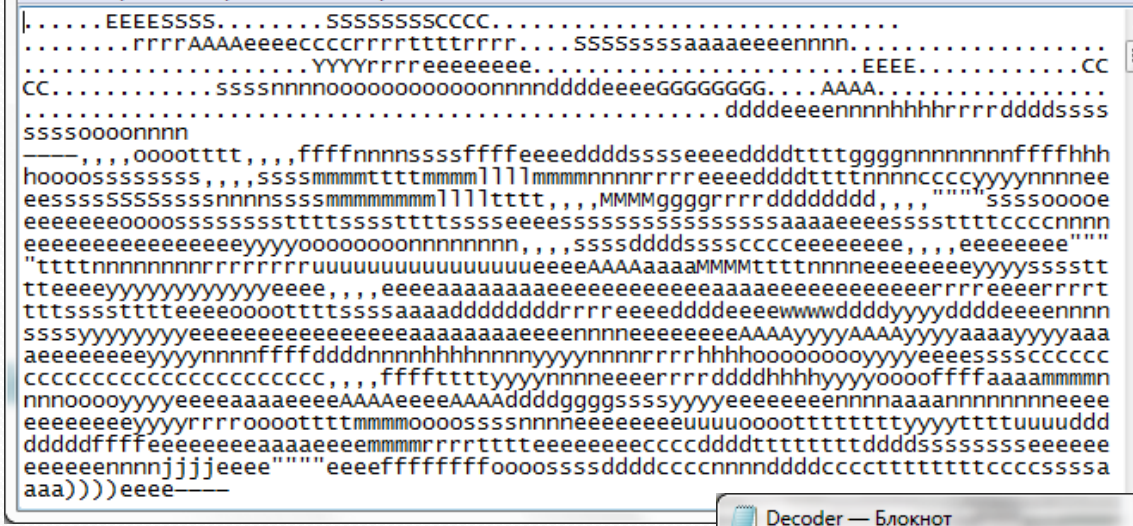
- Кодування – кодує файли за допомогою алгоритма Burrows–Wheeler Transform(BWT).
- Декодування – Декодує файли в вихідне положення
- Стиснути - виконання стиснення за допомогою BWT + RLE.
- Розтиснути – розтиснення файла за допомогою алгоритмів BWT + RLE.
- ? – Виклик довідки.

Процес кодування, архівування

Порівняння кодування

Encoder — Блокнот

Файл Правка Формат Вид Справка



← Кодований файл

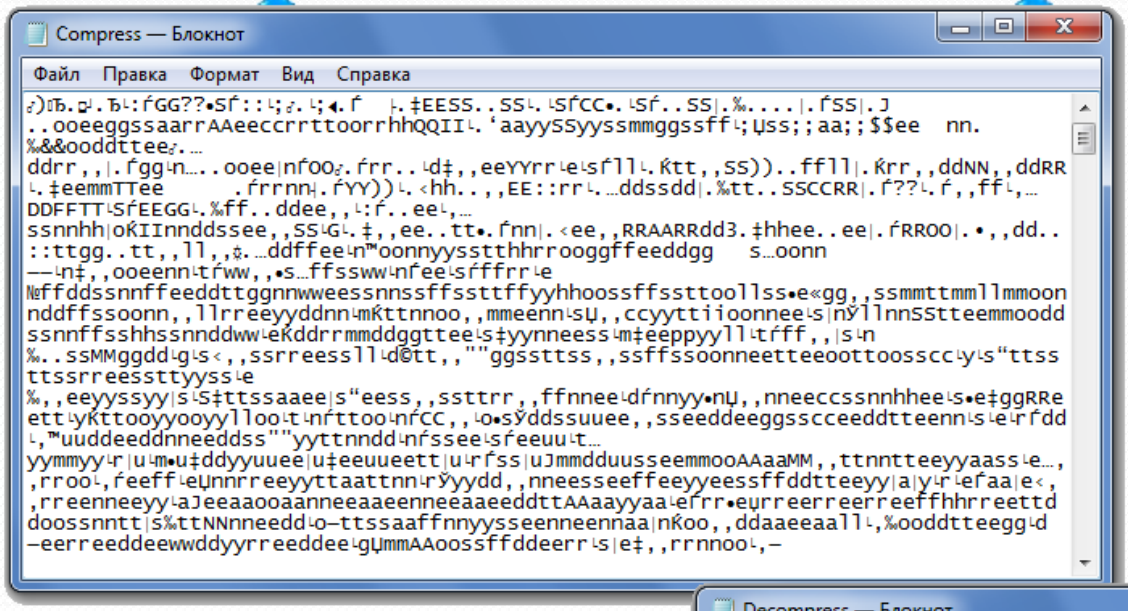
Вихідний файл->

Decoder — Блокнот

Файл Правка Формат Вид Справка

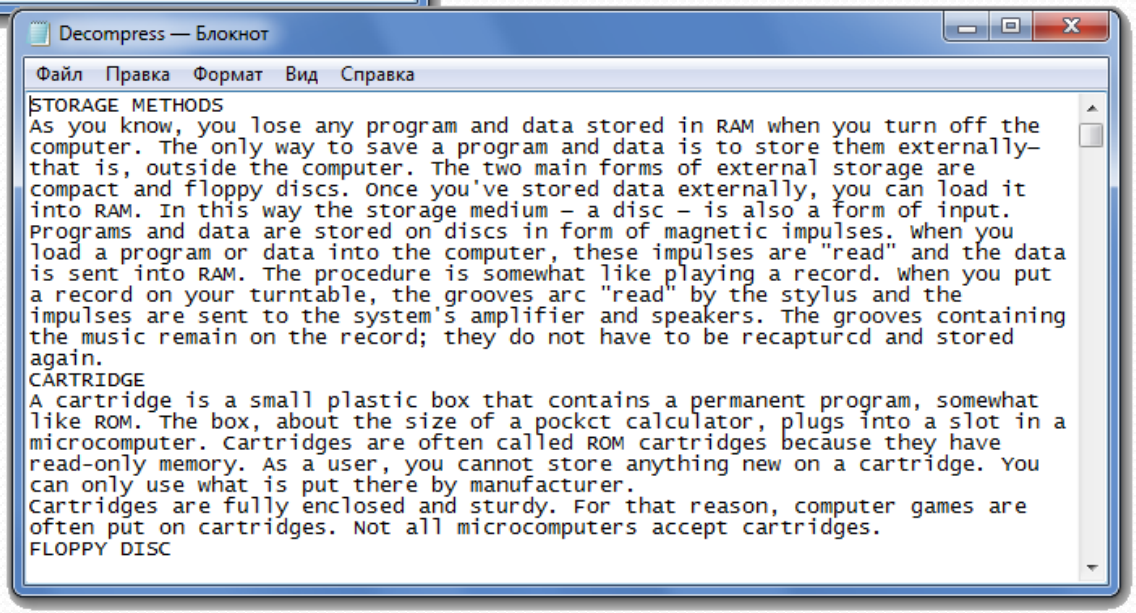
STORAGE METHODS
As you know, you lose any program and data stored in RAM when you turn off the computer. The only way to save a program and data is to store them externally–that is, outside the computer. The two main forms of external storage are compact and floppy discs. Once you've stored data externally, you can load it into RAM. In this way the storage medium – a disc – is also a form of input. Programs and data are stored on discs in form of magnetic impulses. When you load a program or data into the computer, these impulses are "read" and the data is sent into RAM. The procedure is somewhat like playing a record. When you put a record on your turntable, the grooves arc "read" by the stylus and the impulses are sent to the system's amplifier and speakers. The grooves containing the music remain on the record; they do not have to be recapturcd and stored again.
CARTRIDGE
A cartridge is a small plastic box that contains a permanent program, somewhat like ROM. The box, about the size of a pockct calculator, plugs into a slot in a microcomputer. Cartridges are often called ROM cartridges because they have read-only memory. As a user, you cannot store anything new on a cartridge. You can only use what is put there by manufacturer.
Cartridges are fully enclosed and sturdy. For that reason, computer games are often put on cartridges. Not all microcomputers accept cartridges.
FLOPPY DISC

Порівняння архівування

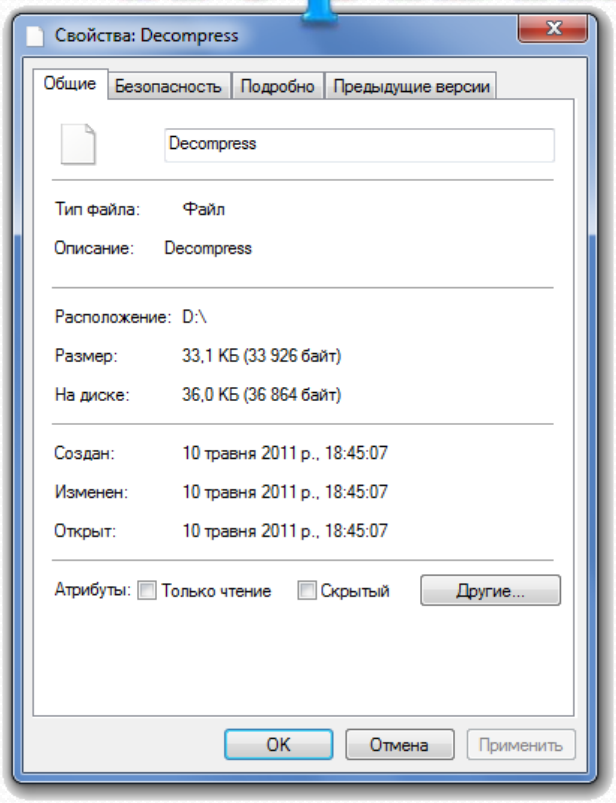


← Стиснутий файл

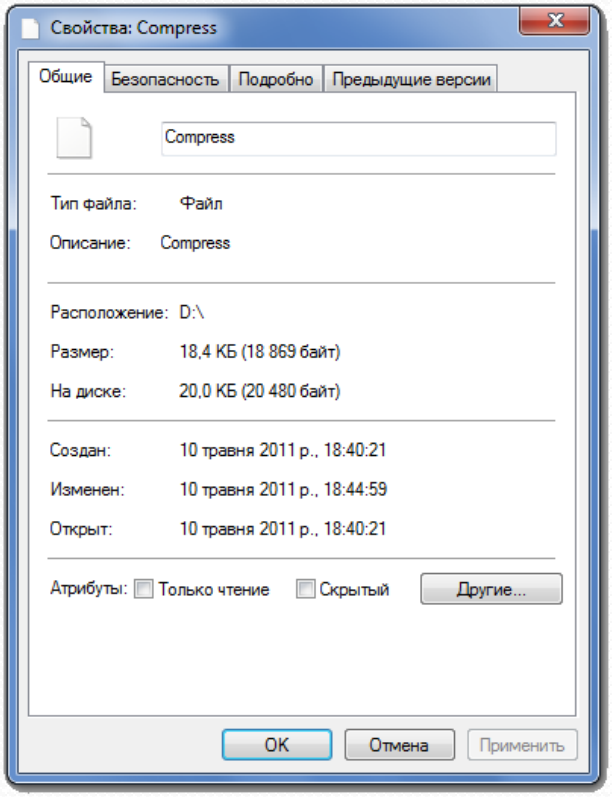
Вихідний файл->



Порівняння розмірів



<-Вихідний файл



Архівований файл->

Дякуємо за увагу!