

Сжатие данных

Материал из Википедии — свободной энциклопедии

Сжатие данных — процедура перекодирования [данных](#), производимая с целью уменьшения их объёма. Применяется для более рационального использования устройств хранения и передачи данных.

Сжатие бывает без потерь (когда возможно восстановление исходных данных без искажений) или с потерями (восстановление возможно с искажениями, несущественными с точки зрения дальнейшего использования восстановленных данных). Сжатие без потерь обычно используется при обработке компьютерных программ и данных, реже — для сокращения объёма [звуковой](#), [фото](#)- и [видеоинформации](#). Сжатие с потерями применяется для сокращения объёма [звуковой](#), [фото](#)- и [видеоинформации](#), оно значительно эффективнее сжатия без потерь.

Сжатие основано на устранении [избыточности информации](#), содержащейся в исходных данных. Примером избыточности является повторение в [тексте](#) фрагментов (например, слов естественного или машинного языка). Подобная избыточность обычно устраняется заменой повторяющейся последовательности более коротким значением (кодом). Другой вид избыточности связан с тем, что некоторые значения в сжимаемых данных встречаются чаще других, при этом возможно заменять часто встречающиеся данные более короткими кодами, а редкие — более длинными (вероятностное сжатие). Сжатие данных, не обладающих свойством избыточности (например, [случайный сигнал](#) или шум), невозможно без потерь. Также, обычно невозможно сжатие зашифрованной информации.

Содержание

- [1 Алгоритмы сжатия текстов/файлов неизвестного формата](#)
- [2 Библиография](#)
- [3 См. также](#)
- [4 Ссылки](#)

Алгоритмы сжатия текстов/файлов неизвестного формата

Имеется 2 основных подхода к сжатию файлов неизвестного формата.

- На каждом шаге [алгоритма сжатия](#) либо следующий символ помещается как есть (со специальным флагом помечающим, что он не сжат), либо указываются границы слова из предыдущего куска, которое совпадает со следующими символами файла. Разархивирование сжатых таким образом файлов выполняется очень быстро, поэтому эти алгоритмы используются для создания самораспаковывающихся программ.
- Для каждой последовательности в каждый момент времени собирается статистика её встречаемости в файле. На основе этой статистики вычисляется вероятность значений для очередного символа. После этого можно применять ту или иную разновидность статистического кодирования, например, [арифметическое](#)

[кодирование](#) или [кодирование Хаффмана](#) для замены часто встречающихся последовательностей на более короткие, а редко встречающихся — на более длинные.

Библиография

- *Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин.* Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. — Диалог-МИФИ, 2002. — С. 384. — [ISBN 5-86404-170-X](#) 3000 экз.
- *Д. Сэломон.* Сжатие данных, изображения и звука. — М.: Техносфера, 2004. — С. 368. — [ISBN 5-94836-027-X](#) 3000 экз.

См. также

- [Сжатие данных с потерями](#)
- [Сжатие без потерь](#)
- [Список алгоритмов](#)

Ссылки

- [Крупнейший в России ресурс, посвященный сжатию](#)

Источник

«http://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»

Категория: [Сжатие данных](#)