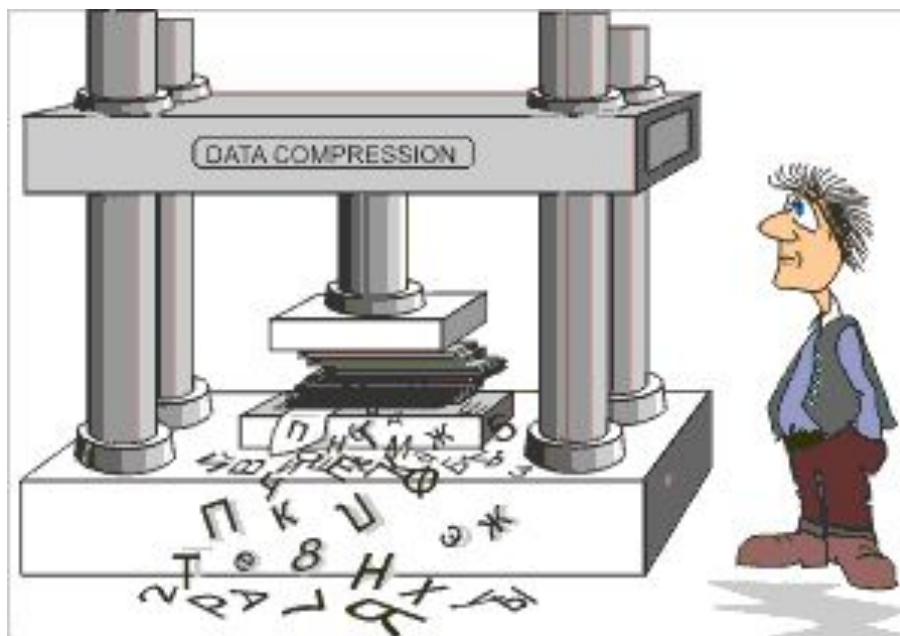


Технологии сжатия данных

Подготовила: Субботина Е. А.

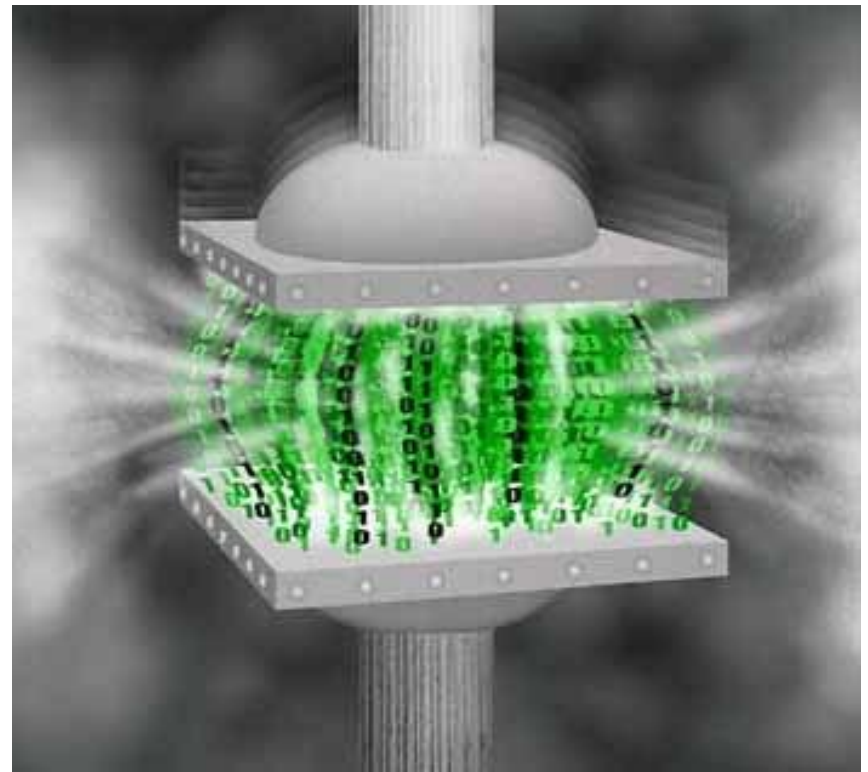
Введение



- **Сжатие данных** ([англ. data compression](#)) — алгоритмическое преобразование [данных](#), производимое с целью уменьшения их объёма.
- Сжатие основано на устранении [избыточности](#)

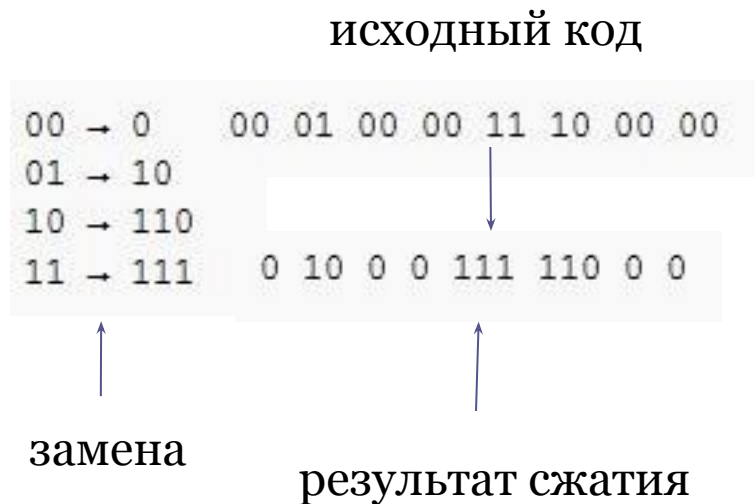
Принципы сжатия данных

- В основе любого способа сжатия лежит модель избыточности.
- Модель избыточности может быть :
 - Статической
 - Либо может строиться или параметризоваться на этапе сжатия (восстановления)
- Все методы делятся на :
 - Сжатие без потерь
 - Сжатие с потерями



Сжатие без потерь

- Пример кодирования двоичных последовательностей



Такая подстановка является префиксным кодом

- Большинство алгоритмов сжатия без потерь работают в две стадии:
 - на первой генерируется *статистическая модель* для входящих данных,
 - вторая отображает входящие данные в битовом представлении, используя модель для получения чаще используемых данных.

Сжатие с потерями

Существуют две основных схемы сжатия с потерями:

- В трансформирующих кодеках
 - фреймы изображений или звука трансформируются в новое базисное пространство
 - производится квантование.
 - Трансформация может осуществляться
 - либо для всего фрейма целиком
 - либо поблочно
 - Результат затем сжимается энтропийными методами.
- В предсказывающих кодеках
 - предыдущие и/или последующие отсчеты данных используются для того, чтобы предсказать текущий отсчет изображения или звука.
 - Ошибка между предсказанными данными и реальными вместе с добавочной информацией, необходимой для производства предсказания, затем квантуется и кодируется.

Анализ методов сжатия графической информации

Для различных типов изображения целесообразно применять подходящие типы алгоритмов сжатия.

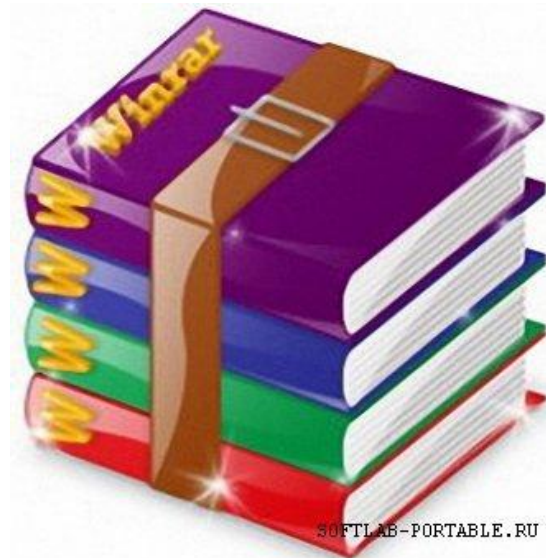
Тип графической информации	Алгоритм сжатия	Графические форматы
Рисунки типа аппликации, содержащие большие области однотонной закрашки.	Заменяет последовательность повторяющихся величин (пикселей одинакового цвета) на две величины (пиксель и количество его повторений).	BMP, PCX
Рисунки типа диаграмм.	Осуществляет поиск повторяющихся в рисунке "узоров".	TIFF, GIF
Отсканированные фотографии, иллюстрации	Основан на том, что человеческий глаз очень чувствителен к изменению яркости отдельных точек изображения, но гораздо хуже замечает изменение цвета. При глубине цвета 24 бита, компьютер обеспечивает воспроизведение более 16 млн. различных цветов, тогда как человек вряд ли способен различить.	JPEG

Примеры



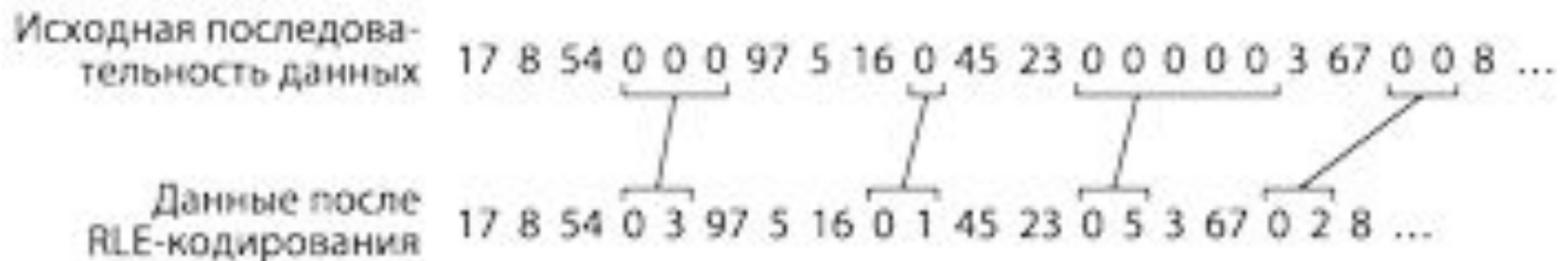
- Форматы сжатия с потерями информации :
 - JPEG - для графических данных;
 - MPG - для видеоданных;
 - MP3 - для аудиоданных.

- Форматы сжатия без потери информации:
 - GIF, TIFF - для графических данных;
 - AVI - для видеоданных;
 - ZIP, ARJ, RAR, CAB, LH - для произвольных типов данных.



Алгоритмы сжатия без потерь

- алгоритм RLE (Run Length Encoding);



- алгоритмы группы KWE(KeyWord Encoding);
- алгоритм Хаффмана.

KWE-кодирование

Пример кодовой таблицы

	Кодовое число	Преобразование
Идентичный код	0000	0
	0001	1
	0254	254
	0255	255
Уникальный код	0256	145 201 4
	0257	243 245
	4095	XXX XXX XXX

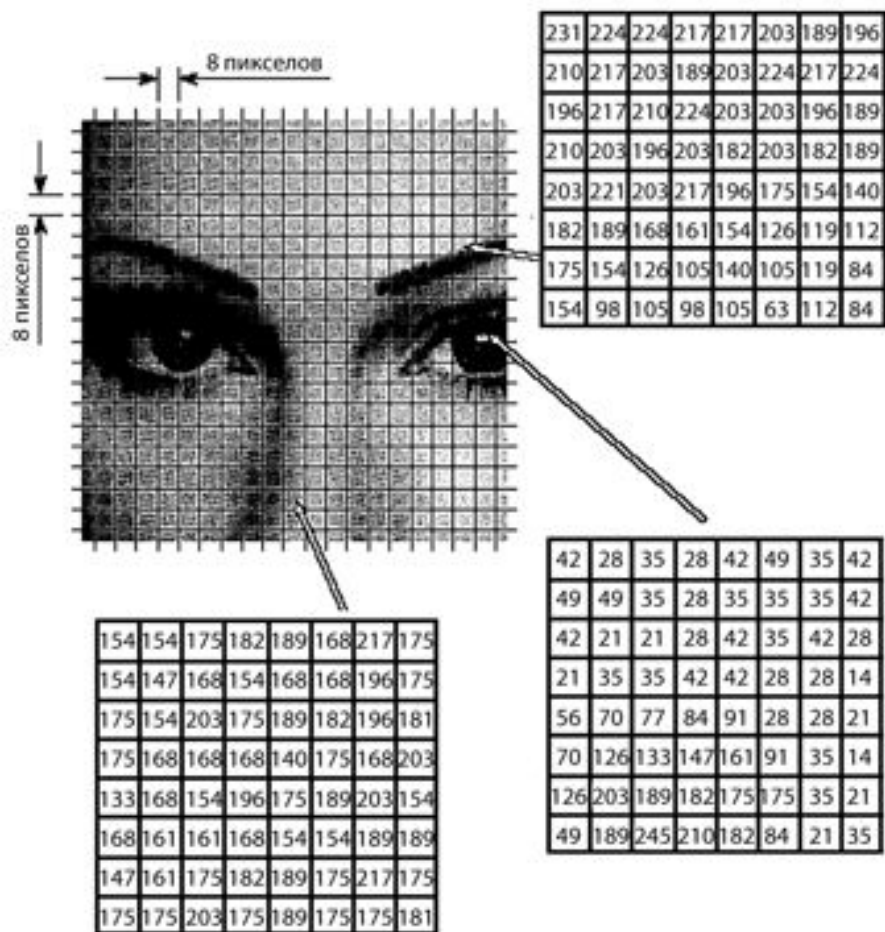
Исходная последовательность данных

123 145 201 4, 119 89 243 245, 59 11 206 145 201 4, 243 245, ...

123 256 119 89 257 59 11 206 256 257...

Кодированные данные

Алгоритм сжатия с потерями-JPEG



Характеристики алгоритмов сжатия и их применимость

- Коэффициент сжатия

$$k = \frac{S_o}{S_c}$$

объём исходных данных

объём сжатых

Чем выше коэффициент сжатия, тем алгоритм эффективнее.

Следует отметить:

- Если $k = 1$, то алгоритм не производит сжатия.
- Если $k < 1$, то алгоритм порождает сообщение большего размера, нежели несжатое.
- Коэффициент сжатия может быть как постоянным, так и переменным.
- Во втором случае он может быть определён либо для каждого конкретного сообщения, либо оценён по некоторым критериям:
 - средний (обычно по некоторому тестовому набору данных);
 - максимальный (случай наилучшего сжатия);
 - минимальный (случай наихудшего сжатия);

Допустимость потерь

- Для некоторых типов данных искажения не допустимы в принципе. В их числе -
 - символические данные
 - жизненно важные данные
 - многократно подвергаемые сжатию и восстановлению промежуточные данные при многоэтапной обработке графических, звуковых и видеоданных.

Алгоритмы сжатия данных неизвестного формата

- Имеется два основных подхода к сжатию данных неизвестного формата:
 - На каждом шаге алгоритма сжатия очередной сжимаемый символ
 - либо помещается в выходной буфер сжимающего кодера как есть,
 - либо группа из нескольких сжимаемых символов заменяется ссылкой на совпадающую с ней группу из уже закодированных символов.
 - Для каждой сжимаемой последовательности символов однократно либо в каждый момент времени
 - собирается статистика её встречаемости в кодируемых данных.
 - На основе этой статистики вычисляется вероятность значения очередного кодируемого символа .
 - После этого применяется та или иная разновидность энтропийного кодирования, например, арифметическое кодирование или кодирование Хаффмана.

ССЫЛКИ

- <http://school497.ru/download/u/02/les15/int.html>
- <http://www.russianelectronics.ru/leader-r/review/8602/doc/46598/>
- <http://www.victoria.lviv.ua/html/informatika/lecture9.htm>
- <http://ru.wikipedia.org/>