

*Д. Ватолин, А. Ратушняк,
М. Смирнов, В. Юкин*

Методы сжатия данных

Предисловие

Содержание книги:

Введение

Раздел 1. Универсальные методы сжатия

Раздел 2. Алгоритмы сжатия изображений

Раздел 3. Алгоритмы сжатия видео

Приложение 1

Приложение 2

**ISBN 5-86404-170-X
2002**

Предисловие

Основная задача, которая ставилась при написании этой книги — изложить в одном издании и достаточно единообразно современные методы сжатия данных. Тема эта необъятная. Все разделы достойны отдельных книг, а развернутое описание методов сжатия видеоданных требует нескольких томов. Поэтому в изложении основной упор делался на базовые идеи и концепции, используемые при сжатии. Нам бы хотелось, чтобы, изучив изложенные в книге методы, читатель мог понять принципы работы большинства компрессоров и разобраться в методах сжатия данных по исходным текстам программ.

Многие в первую очередь зададут вопрос: «Где взять исходные тексты архиваторов? И чтобы степень сжатия была хорошая!». Это не проблема. Сейчас в любом достаточно крупном городе России можно без труда приобрести компакт-диски с дистрибутивом Linux. В составе исходных текстов этой операционной системы есть текст на языке программирования C, позволяющий работать со сжатыми дисками файловой системы NTFS. Следовательно, можно посмотреть, как организовано сжатие в операционных системах типа Windows NT. Также в состав практически всех дистрибутивов Linux входят исходные тексты утилит сжатия `gzip` и `bzip2`, дающих неплохие результаты. Поэтому исходные тексты обычных программ сжатия, имеющих хорошие характеристики, можно найти буквально на каждом лотке с CD-ROM и на десятках тысяч сайтов с дистрибутивами Linux в Интернет. Кроме того, есть сайты с исходными текстами эффективных и качественно реализованных архиваторов, которые пока менее известны и распространены. Например, сайт проекта 7-Zip <http://www.7-zip.org>. Примечательно, что такая ситуация складывается не только со сравнительно простыми универсальными архиваторами (10-200Кб исходных текстов), но и с такими сложными системами, как видеокодеки (400-4000Кб текста). Свободно доступны тексты кодеков вполне промышленного качества, таких как OpenDivX и On2, тексты конвертера VirtualDub, поддерживающего несколько фор-

матов видео. Таким образом, у человека, имеющего доступ к Интернету и умеющего искать (попробуйте начать с <http://dogma.net/DataCompression/>), возникает проблема не найти исходные тексты программ, а понять, *как* и *почему* эти программы работают. Именно в этом мы и пытаемся помочь нашим читателям.

Мы старались не только охватить все основные подходы, используемые при сжатии разных типов данных, но и рассказать о современных специальных техниках и модификациях алгоритмов, зачастую малоизвестных. Естественно, полнее всего раскрыты темы, в области которых авторы наиболее компетентны. Поэтому некоторые алгоритмы рассмотрены глубоко и основательно, в то время как другие, также достойные, описаны конспективно или лишь упомянуты. Тем не менее, главная цель достигнута: изложение получилось достаточно интересным, практическим и охватывающим ключевые методы сжатия данных различных типов.

В тексте книги встречается большое количество русских имен. Для многих будет неожиданностью, что очень много интересных и эффективных компрессоров написано программистами, живущими в России и других странах СНГ. В частности, это RAR (автор Евгений Рошал), 7-Zip, VIX, UFA, 777 (Игорь Павлов), BMF, PPMd, PPMonstr (Дмитрий Шкарин), YBS (Вадим Юкин), ARI, ERI (Александр Ратушняк), PPMN (Максим Смирнов), PPMY (Евгений Шелвин), и многие другие. Причем по данным тестирования, например, на сжатие исполняемых файлов (<http://compression.ca/act-executable.html>) на 31.01.2002 в рейтинге, содержащем 154 архиватора, программы наших соотечественников занимают 8 позиций в двадцатке лучших. Авторам этой книги искренне хотелось бы, чтобы наших архиваторов в таких рейтингах становилось все больше.

Написание материалов книги и ответственность были распределены следующим образом:

Д. Ватолиным написан пункт «Арифметическое сжатие. Классический вариант алгоритма», раздел 2 за исключением пункта «Методы обхода плоскости», весь раздел 3, а также приложение 2;

А. Ратушняком — пункты «Разделение мантисс и экспонент», «Нумерирующее кодирование», «Векторное квантование», «Методы обхода плоскости», главы «Кодирование источников данных типа "аналоговый сигнал"», «Обобщенные методы сортирующих преобразований», а также части введения «Определения. Аббревиатуры и классификации методов сжатия», «Замечание о методах, алгоритмах и программах»;

М. Смирновым — главы «Словарные методы сжатия данных», «Методы контекстного моделирования», пункт «Препроцессинг текстов» и приложение 1;

В. Юкиным — глава «Преобразование Барроуза-Уилера», пункты «Интервальное кодирование», «Препроцессинг нетекстовых данных», «Выбор метода».

Часть введения «Сравнение алгоритмов по степени сжатия» написана совместно А. Ратушняком и М. Смирновым.

В заключение мы хотим выразить свою благодарность Дмитрию Шкарину, Евгению Шелвину, Александру Жиркову, Владимиру Вежневцу, Алексею Игнатенко: их конструктивная критика, интересные дополнения, замечания и советы способствовали существенному улучшению качества книги. Также мы благодарны лаборатории машинной графики ВМиК МГУ и персонально Юрию Матвеевичу Баяковскому и Евгению Викторовичу Шикину за организационную и техническую поддержку.

Всю ответственность за все ошибки и неточности, слабость изложения отдельных пунктов и прочие недостатки книги несут только авторы. Мы будем признательны читателям за их отзывы и критические замечания, отправленные непосредственно нам по электронной почте на адрес compression@graphicon.ru.

Исходные тексты программ, ответы на контрольные вопросы и упражнения вы сможете получить на сайте <http://compression.graphicon.ru/>.

*Дмитрий Ватолин, Александр Ратушняк,
Максим Смирнов, Вадим Юкин
Москва – Новосибирск – Санкт-Петербург, 2001 год*