

Сжатие и архивация данных

Общеизвестна достаточно высокая стоимость хранения данных на магнитных носителях и особенно передачи их по линиям связи, поэтому существует необходимость, по-возможности, уменьшать их объем, разумеется, не жертвуя при этом их качеством.

В принципе любая сумма данных, циркулирующих в медицине, является в той или иной степени избыточной (и очень редко недостаточной). Степень этой избыточности зависит от вида данных. Так, видеofilm, получаемый, например, при рентгеноконтрастном исследовании сердца, или аускультативный звуковой ряд имеют степень избыточности в несколько раз большую, чем графические данные (рентгенограмма контрастированного сердца при том же исследовании). В свою очередь, избыточность графических данных значительно превышает таковую у текстовых данных. Интересен тот факт, что русский текст более избыточен, чем, например, английский или немецкий.

В медицинской практике избыточность данных используется для повышения качества информации. На этом принципе строятся многие статистические выкладки. Увеличение количества данных улучшает качество восприятия их человеком, например, видеofilm или звука. Одна при хранении и передаче данных по каналам связи величина их объема приобретает критическое значение. Поэтому в информационных технологиях широкое распространение получили *методы сжатия данных*. Близкий по содержанию к нему термин – *архивация данных*. Осуществляют эти операции с данными с помощью программ, называемых *архиваторами*.

Выделяют два варианта сжатия (архивации):

1. *Сжатие (архивация) файлов и папок*, которое применяется при подготовке данных к хранению, особенно на переносных магнитных носителях, и транспортировке их по электронным каналам связи.
2. *Сжатие (архивация) дисков* используется в целях увеличения их рабочего пространства.

Обе эти процедуры могут выполняться двумя способами: 1) *с частичной потерей данных* при последующем разархивировании папок и файлов, 2) с полным восстановлением содержания данных при разархивировании. Первый способ позволяет сжать исходную совокупность данных в десятки раз. Он применяется при работе с данными, для которых потеря их части существенно

не отражается на их использовании и восприятии: медицинские изображения, рисунки, мультимедийные комплексы, звукозапись. Следует, однако, отметить, что сжатие медицинских изображений (рентгенограмм, томограмм и др.) более чем в 5 раз приводит к некоторым искажениям, которые нужно учитывать в практической работе медицинского учреждения. Второй способ более щадящий: совокупность данных сжимается всего в несколько раз, иногда всего в 1,5–2 раза. Это сжатие применяется для тех данных, при разархивировании которых утрата даже небольшой их доли существенно отражается на их содержании. К таким данным относятся все текстовые документы, базы данных, программы.

Основными программами для сжатия данных с потерей являются: для графических данных – *JPG*, для видеофильмов – *MPG*, для звукозаписи – *MP3*. Характерными программами (точнее форматами) для сжатия данных без их потери при разархивировании являются: для графических данных – *GIF*, *TIF*, *PCX*, *DjVu*, для видеофильмов – *AVI*, для любых типов данных – *ARJ*, *ZIP*, *RAR*. При разархивировании (распаковке) данных применяют программы *PKUNZIP.EXE*, *UNRAR.EXE* и некоторые другие.

Следует учесть, что если сжатые данные передаются другому лицу – по каналам связи или на внешних носителях – у получателя этих данных должны быть программы по адекватному их разархивированию. Если таковой у получателя нет, отправляющий должен прикрепить к отправляемому архиву короткий специальный программный модуль. Такие архивы данных носят названия *самораспаковывающихся*. Их наименование заканчивается расширением *.EXE*.

Необходимо отметить, что современное программное обеспечение позволяет не только осуществлять архивацию и разархивацию файлов и папок, но помогает также выполнять ряд дополнительных функций. С его помощью можно осуществлять тестирование архивов, восстановление поврежденных архивов, извлечение одного или нескольких требующихся файлов из архива и добавление в него новых, создание распределенных архивов на нескольких переносных носителях или на жестком диске (например, программы *.WINZIP*, *.WINRAR*). Такие программные средства носят название *диспетчеров архивов*.