

СЖАТИЕ И АРХИВИРОВАНИЕ ДАНЫХ

**Учебно-методическое пособие
для аспирантов**

Минск БГТУ 2004

Учреждение образования
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ»

СЖАТИЕ И АРХИВИРОВАНИЕ ДАНЫХ

**Учебно-методическое пособие
для аспирантов**

Минск 2004

УДК 681.3.05
ББК 32.811.4
С 33

Рассмотрено и рекомендовано к изданию редакционно-издательским советом университета

Составители:

Н.В. Пацей, П.П. Урбанович

Рецензенты:

ректор ИИТ БГУИР доцент А.Г. Якубенко;
проректор БГУИР профессор С.П. Кундас;
зав. кафедрой АППиЭ БГТУ доцент И.Ф. Кузьмицкий

Сжатие и архивирование данных. Учеб.-метод. пособие для
С 33 аспирантов / Сост. Н.В. Пацей, П.П. Урбанович. – Мн.:
БГТУ, 2004. – 32 с.

ISBN 985-434-333-2

Учебно-методическое пособие содержит сведения, необходимые для выполнения двух лабораторных работ аспирантами. В пособии изложены основные алгоритмы сжатия, рассмотрены функции и команды программ-архиваторов WinZip, WinRAR, WinArj32 и Windows Backup.

УДК 681.3.05
ББК 32.811.4

ISBN 985-434-333-2

© Учреждение образования
«Белорусский государственный
технологический университет», 2004

ВВЕДЕНИЕ

Вопросы сжатия данных достаточно остро стоят в различных областях науки и техники, везде, где требуется хранение и передача информации. Во-первых, это связано со все еще достаточно высокой стоимостью носителей информации (магнитные и оптические диски, ПЗУ, ОЗУ и т.д.), во-вторых, с необходимостью передачи больших потоков информации по перегруженным линиям связи (радио- и оптическая связь, телефония, сети).

Методы сжатия увеличивают энтропию сообщения, то есть уменьшают длину файла, сохраняя при этом всю информацию (при использовании методов сжатия без потерь). Часто в файлах некоторые значения байта встречаются чаще других. Таким образом, за счет использования для каждого значения байта кода различной длины можно значительно уменьшить общий размер данных. Эта базовая идея лежит в основе алгоритмов сжатия.

Основными понятиями, которые используют при рассмотрении проблемы сжатия данных, являются понятия «архивирование», «техника сжатия», «коэффициент сжатия».

Техника сжатия (компрессии, упаковки) данных основана на замене повторяющихся последовательностей байтов более короткими кодами и составлении таблицы соответствия, используемой в последующем для восстановления данных в первоначальном виде, т.е. для их распаковки (декомпрессии, разжатия). Таким образом, архив логически делится на 2 компонента – таблицу соответствия и собственно сжатые данные. **Архивирование** – упаковка (сжатие) файла или группы файлов с целью уменьшить место, занимаемое им на диске. **Коэффициент сжатия** есть отношение размера сжатого файла к ее первоначальному размеру (наиболее часто используемое соотношение).

Существует два основных способа сжатия: статистический и словарный. В статистическом сжатии каждому символу присваивается код, основанный на вероятности его появления в тексте. Высоковолевым символам (наиболее часто встречающимся в сообщениях) приписываются короткие коды и наоборот. В словарном методе группы последовательных символов или «фраз» заменяются кодом. Замененная фраза может быть найдена в некотором, ранее сформированном «словаре».

1. АЛГОРИТМЫ СЖАТИЯ ТЕКСТОВОЙ ИНФОРМАЦИИ

1.1. Код Шеннона – Фано

Алгоритм Шеннона – Фано состоит из следующих шагов:

- 1) сортировка входных символов по вероятности их встречаемости в убывающем порядке;
- 2) разделение всего ряда символов на две части таким образом, чтобы каждая часть имела равные или приблизительно равные суммы вероятностей;
- 3) кодирование символов первой части битом *ноль* (один) и символов второй части – битом *один* (ноль);
- 4) повторение шага 2 и рекурсивное продолжение процесса для каждой из частей до тех пор, пока каждая из них не будет содержать по одному символу.

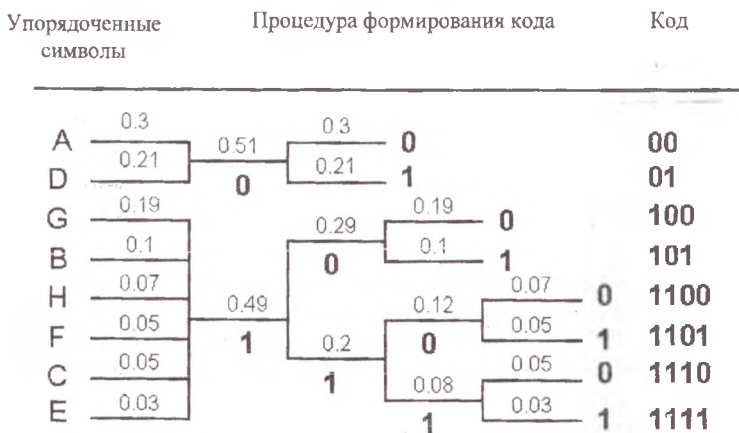


Рис. 1. Пример кодирования данных алгоритмом Шеннона – Фано

1.2. Код Хаффмана

Код Хаффмана включает в себя большой класс кодов переменной длины и однозначного декодирования. Алгоритм состоит из следующих шагов:

- 1) сортировка ссылок (узлов) входных символов (не изменяя ме-

стонахождения каждого из символов) по вероятности их встречаемости в убывающем порядке;

2) объединение двух символов с наименьшими вероятностями в композицию символов с вероятностью, равной сумме исходных вероятностей;

3) повторение шага 2 до тех пор, пока не получится композиция с вероятностью 1 (называемая корнем). Полученная структура называется деревом Хаффмана;

4) проход по дереву от корня до соответствующего символа и присвоение 0 (1) левой и 1(0) – правой ветви.

Рассмотрим пример. Пусть имеется файл, основанный на использовании 6 различных символов (говорят – мощность используемого алфавита равна 6). Подсчитываем вероятность (частоту) вхождения каждого из символов в файл и получаем табл. 1.

Таблица 1

Символ	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
Вероятность	0.1	0.2	0.3	0.05	0.25	0.1

Выбираем из таблицы символы с наименьшей вероятностью. В нашем случае это *D* (0.05) и какой либо символ из *F* или *A* (0.1); можно взять любой из них, например, *A*. Формируем из «узлу» *D* и *A* композицию (новый «узел»), вероятность для которого будет равна сумме вероятностей *D* и *A*: $0.05+0.1=0.15$ (рис. 2). Снова ищем два символа с самыми низкими вероятностями. Самая низкая вероятность теперь соответствует *F* и новому «узлу» (*D* и *A*). Снова выполняем операцию слияния «узлу»: $0.1+0.15=0.25$. Продолжаем этот процесс до тех пор, пока все не сведется к одному «узлу». После формирования дерева (рис. 2) кодируем файл.

Кодирование всегда начинается с корня. Кодирова символ, прослеживаем вверх по дереву все повороты ветвей. Если делается левый поворот, запоминается 0-й бит, и аналогично – 1-й бит для правого поворота. Получаем табл. 2.

Таблица 2

Символ	<i>C</i>	<i>A</i>	<i>D</i>	<i>F</i>	<i>B</i>	<i>E</i>
Код	00	0100	0101	011	10	11
Размер (бит)	2	4	4	3	2	2

Таким образом, каждый символ изначально представлялся,

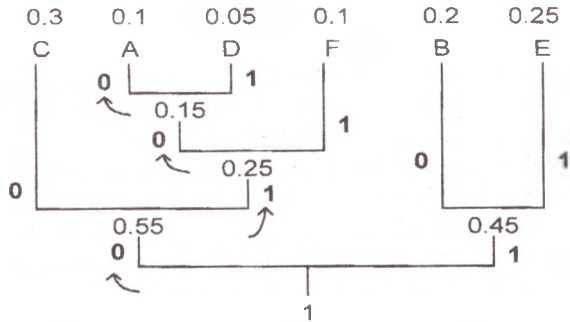


Рис. 2. Дерево кодирования по алгоритму Хаффмана

положим, 8-ю битами (минимально необходимое количество битов, соответствующих одному символу алфавита, есть наименьшее целое число, большее или равное двоичному логарифму числа, составляющего мощность алфавита). В процессе кодирования число битов, необходимых для представления каждого символа, уменьшилось.

Для восстановления первоначального файла необходимо иметь декодирующее дерево. Следовательно, дерево должно храниться вместе со сжатым текстом. Это превращается в итоге в увеличение размеров выходного файла.

Алгоритм сжатия Хаффмана, основанный на статистических колебаниях распределения некоторых значений байтов, лег в основу нескольких эффективных методов сжатия, известных как арифметическое кодирование (arithmetic coding), энтропийное кодирование (entropy coding) или Q-кодирование (Q-coding).

1.3. Арифметические коды

Арифметическое кодирование – метод кодирования данных в вещественных числах на интервале $[0; 1)$. Рассмотрим пример:

1) определим символы источника как $\{00, 01, 10, 11\}$ и соответственно вероятности этих символов: $\{0.1; 0.4; 0.2; 0.3\}$;

2) разделим интервал $[0; 1)$ на 4 подынтервала вида $[x, y)$ (включающий x и исключаящий y), а именно: $[0; 0.1)$, $[0.1; 0.5)$, $[0.5; 0.7)$, $[0.7; 1)$;

3) полученные данные могут быть сведены в табл. 3;

4) для кодирования бинарной последовательности, например вид $10\ 00\ 11\ 00\ 10\ 11\ 01$, выбираем первый символ 10 и находим его кодо-

вый диапазон $[0.5; 0.7)$. Так как диапазон второго символа $00 - [0; 0.1)$, кодируем путем выбора первого диапазона 10 интервала, получаем новый интервал $[0.5; 0.52)$. Аналогично для кодирования третьего символа 11 мы получаем новый интервал $[0.51; 0.52)$ и т.д. (рис. 3). Сжатый выход этого сообщения может быть любым числом из последнего интервала. В данном примере выбираем число 0.51439 .

Таблица 3

Символы	00	01	10	11
Вероятности	0.1	0.4	0.2	0.3
Кодовые интервалы	$[0; 0.1)$	$[0.1; 0.5)$	$[0.5; 0.7)$	$[0.7; 1)$

При декодировании полученное декодером число 0.51439 принадлежит интервалу $[0.5; 0.7)$, поэтому первый входной символ – 10 . Пересчет дает результат $(0.51439 - 0.5)/0.5 = 0.02878$, который принадлежит диапазону $[0; 0.1) \rightarrow 00$. Аналогично $(0.1 - 0.02878)/0.1 = 0.7122 \in [0.7, 1) \rightarrow 11$ и т.д.

Предполагается, что и кодеру, и декодеру «известна» длина сообщения, в противном случае процесс декодирования будет продолжаться бесконечно. Арифметическое кодирование чувствительно к ошибкам.

1.4. Словарные алгоритмы

Эти алгоритмы сжатия используют свойство – содержать повторяющиеся кодовые последовательности. Все словарные алгоритмы делятся на две основные группы. Алгоритмы первой группы «стараяются» обнаружить факт того, встречалась ли уже сжимаемая последовательность символов во входных данных и затем вместо ее повторения, в выходных данных, устанавливается на нее указатель (рис. 4 а). Алгоритмы второй группы создают словари из фраз, встречаемых во входных данных. В кодированной последовательности указывается лишь номер фразы в словаре (рис. 4 б).

1.5. LZ алгоритм

Алгоритм LZ (Лемпела – Зива) основан на идее формирования «словаря» общих последовательностей данных. При этом сжатие данных осуществляется за счет замены последовательностей соответствующими кодами из словаря. Так были созданы алгоритмы сжатия LZ77 и LZ78, LZSS и др.

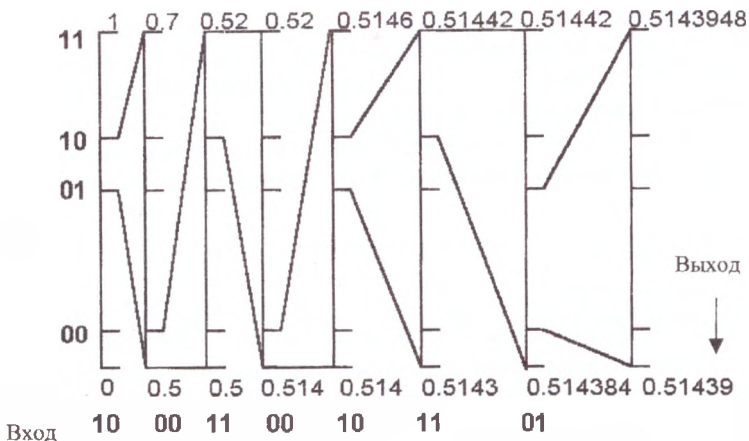


Рис. 3. Пример процедуры арифметического кодирования

Декодирование сжатого текста осуществляется напрямую - происходит простая замена указателя готовой последовательностью из словаря, на которую тот указывает. Одной из форм такого указателя есть пара (i, j) , которая заменяет последовательность из j символов, начинающуюся со смещения i во входном потоке.

По мере выполнения обработки словарь скользит по входному потоку данных. Скользящее окно имеет длину N и состоит из двух частей: последовательности уже закодированных символов, которая и является словарем, и упреждающего буфера, или буфера предварительного просмотра. Идея алгоритма состоит в поиске самого длинного совпадения между строкой буфера и всеми фразами словаря.

Рассмотрим пример сжатия текста «Полили ли лилию» методом LZSS. Обозначим через i смещение, j - длину соответствия или совпадения, s - выходной символ, f - флаг. Пусть длина буфера равна 7 символам, а размер словаря больше длины сжимаемой строки.

Символ переписывается в явном виде, если текущая длина максимального совпадения буфера и какой-то фразы словаря меньше или равна 1. Если записывается символ, то перед ним выдается флаг 0, если указатель - то 1 (табл. 4). Если имеется несколько совпадающих фраз одинаковой длины, то выбирается ближайшая к буферу.

Таким образом, для кодирования текста по алгоритму LZSS по-

требовалось 9 шагов: 6 раз символы передавались в явном виде и 3 раза использовали указатели.

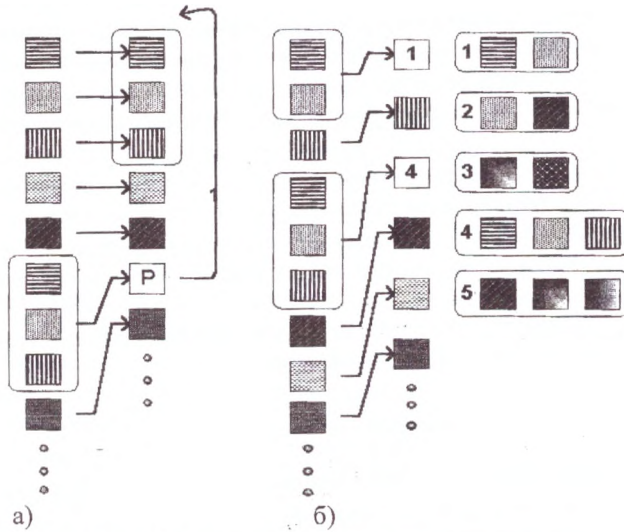


Рис. 4. Примеры словарных алгоритмов сжатия

Таблица 4

Шаг	Скользящее окно		Совпадающая фраза	Закодированные данные			
	Словарь	Буфер		<i>f</i>	<i>i</i>	<i>j</i>	<i>s</i>
1	-	полили_	-	0	-	-	'п'
2	п	олили_л	-	0	-	-	'о'
3	по	лили_ли	-	0	-	-	'л'
4	пол	или_ли_	-	0	-	-	'и'
5	поли	ли_ли_л	ли	1	2	2	-
6	полили	_ли_лил	-	0	-	-	'_'
7	полили_	ли_лили	ли_	1	3	3	-
8	полили_ли_	лилию	лили	1	8	4	-
9	полили_ли_лили	ю	-	0	-	-	'ю'

1.6. Задание №1

1. Изучить принципы приведенных алгоритмов сжатия данных.
2. Используя заданные в табл. 5 исходные тексты, построить де-рево (схему) кодирования. Закодировать (сжать) первые два слова текста одним из методов: Шеннона – Фано, Хаффмана или LZ.
3. Сравнить размеры сжатой и исходной фраз, рассчитать коэффициент сжатия. Размер буквы в исходном тексте считать равным 8 бит.
4. Восстановить сжатый текст.
5. Конвертировать последнее слово исходного текста в табл. 4 в двоичный ASCII код и представить его в арифметическом коде.
6. Найти оптимальный метод сжатия исходного сообщения из табл. 5.
7. Доказать, что для любого входного текста существует несколько различных кодов Хаффмана.

Таблица 5

№	Исходный текст	Метод сжатия
1	Принимайтесь за архивирование сегодня же	Хаффмана
2	Сжатие – есть форма кодирования	Хаффмана
3	Архивировать нужно все данные	Хаффмана
4	Архивируйте данные так часто, как не бойтесь их потерять	Шеннона – Фанно
5	Периодичность архивирования зависит от скорости изменения данных	Хаффмана
6	Ежедневно запускайте архивирование оперативных данных	Шеннона – Фанно
7	Запускайте архивирование менее критичных данных раз в месяц	Шеннона – Фанно
8	Храните архивы на другом компьютере или ZIP-дисках	Шеннона – Фанно
9	Периодически сбрасывайте архив на диски	Хаффмана
10	До города дорога в гору	LZ
11	Проворонила ворона вороненка	LZ
12	Кот ломом колот слона	LZ
13	В такси таксе тесно	LZSS
14	В печи калачи печены	LZSS

2. СОЗДАНИЕ АРХИВОВ

Критериев, позволяющих однозначно предпочесть один архиватор другому, не существует.

Основные характеристики архиваторов:

- 1) коэффициент сжатия;
- 2) скорость сжатия;
- 3) сервис (т.е. набор функций архиватора).

Самыми популярными форматами, в которых хранятся сжатые заархивированные данные, являются ARJ, ZIP, RAR и некоторые другие.

2.1. Архиватор WinZip

Интерфейс. Заголовок содержит название программы – WinZip и имя архива, с которым вы работаете. Панель инструментов позволяет быстро выполнять самые распространенные задачи (рис. 5). Основное окно содержит информацию о файлах в архиве. В строке состояния отображается число и размер выбранных файлов, наряду с другой информацией об архиве (рис. 5).

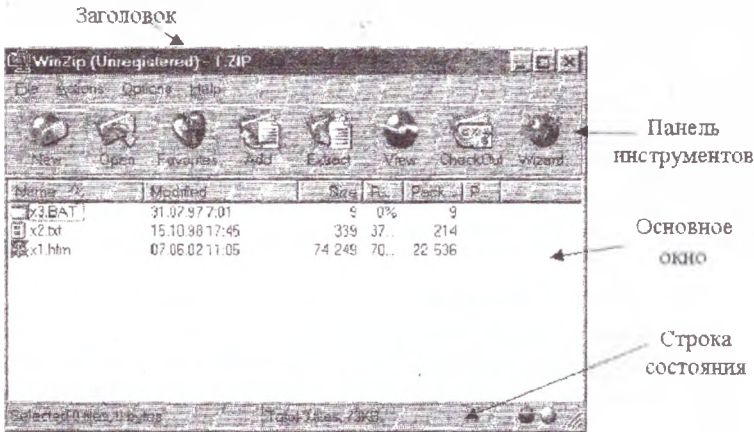


Рис. 5. Основные окна архиватора WinZip

Для создания новых архивов служит команда **File/New Archive** или кнопка **New** на панели инструментов. В диалоговом окне *New Archive* укажите название создаваемого архива, (но не названия

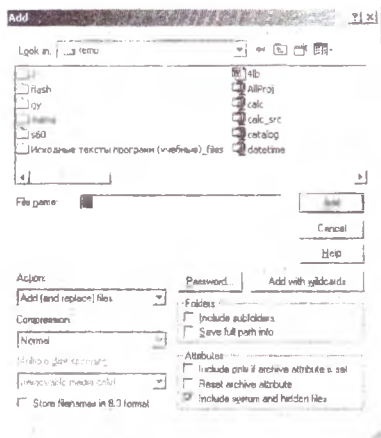


Рис. 6. Диалоговое окно Add

файлов, которые вы хотите архивировать).

Для автоматического вызова диалогового окна *Add* (рис. 6) установите флаг *Add dialog*.

Чтобы добавить файлы к архиву, выполните команду **Actions/Add** или нажмите кнопку **Add** на панели инструментов.

Для добавления файлов с помощью маски в поле *File name* задается маска и нажимается кнопка **Add with wildcards**.

Файлы добавляются также перетаскиванием в окно *WinZip*, значок **WinZip** на рабочем столе,

Zip-файл в *Моем компьютере* или *Проводнике*, значок *Zip*-файла или командой **Add To Zip** контекстного меню.

Для открытия архивов используйте команду **File/Open Archive** или кнопку **Open**. Архив открывается перетаскиванием его в окно *WinZip* или двойным щелчком по названию архива.

Самораспаковывающийся *zip*-файл – это исполнимая программа (*exe*-файл), которая содержит одновременно и *zip*-архив и программу для извлечения его содержимого. **Самораспаковывающиеся архивы создаются** командой **Actions/Make .Exe File**, которая преобразовывает открытый архив в самораспаковывающийся. Если необходимо разбить самораспаковывающийся архив по дискетам, выберите опцию *Safe spanning method*.

2.2. Архиватор WinRAR

Программа-архиватор **WinRAR** поддерживает режимы управления файлами и архивами. Для входа в режим управления файлами запустите **WinRAR**. Для входа в режим управления архивами откройте архив.

Для создания архивов, добавления файлов служит команда **Команды/Добавить файлы в архив**, кнопка **Добавить** или комбинация клавиш *Alt-A*. Выберите имя архива и параметры (рис. 7).

Чтобы создать самораспаковывающийся архив или *SFX*-

архив (от англ. Self-eXtracting) включите опцию *Создать SFX-архив* в диалоге ввода имени архива и его параметров (рис. 7), или войдите в существующий архив и нажмите кнопку **SFX**.

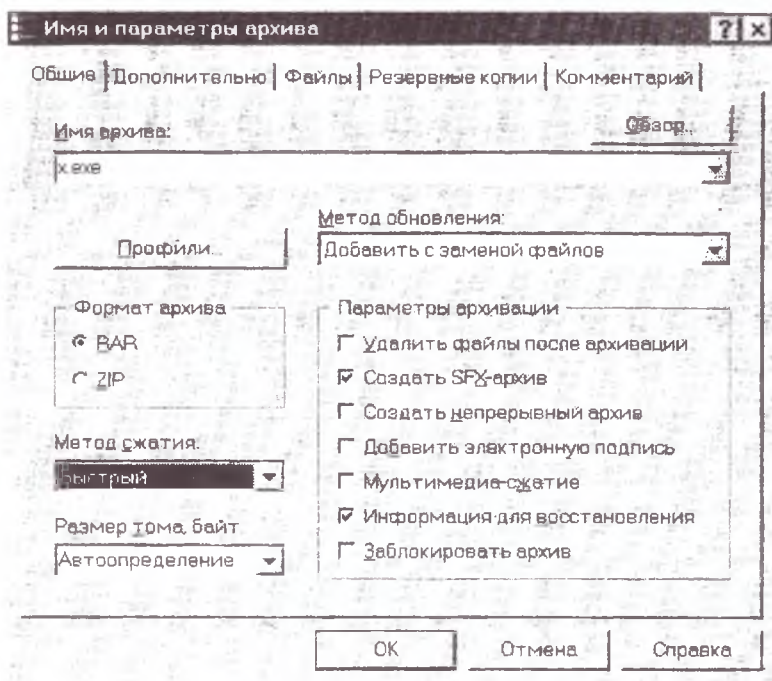


Рис. 7. Диалоговое окно добавления файлов

Для создания самораспаковывающегося архива **GUI SFX-модуля**, работающего в графическом режиме *GUI*, необходимо ввести команды для организации простых процедур установки. Эти команды должны находиться в комментарии *SFX*-архива, доступного через **Команды/Добавить архивный комментарий** или кнопку **Коммент**:

Text=<строка> 'Добавляет <строку> к окну вывода текста модуля *SFX*. Первая команда **Text** сбрасывает стандартное содержимое этого окна перед добавлением <строки>. Следующие команды добавляют <строку> к уже существующему тексту.

Title=<заголовок> 'Устанавливает заголовок окна *SFX*.

Path=<путь> 'Устанавливает путь по умолчанию. Если указан полный путь, начинающийся с '<буквы диска>:', то он используется без изменений. Если <путь> содержит только имя, то он будет рассматриваться как начинающийся от папки *Program Files*. Если вы хотите отменить подобное поведение, начинайте <путь> с символов '\.'

Silent 'Начать извлечение без вывода начального диалога. Файлы будут извлечены в папку, указанную в команде **Path** или в текущую папку, если путь по умолчанию не задан.

Overwrite=[n] 'Если 'n' — 0, то при перезаписи существующих файлов будет выведен запрос на подтверждение. Это действие выполняется по умолчанию, поэтому для него необязательно указывать команду **Overwrite**. Если 'n' отсутствует или равно '1', то все файлы будут перезаписаны без подтверждения. Если 'n' равно '2', то не будет производиться перезапись существующих файлов.

Delete=<имя файла> 'Перед началом извлечения удаляет указанный файл в папке назначения. Допустимы несколько команд **Delete**.

Setup=<программа> '*SFX* попытается запустить <программу> после удачного извлечения файлов. При этом текущей папкой будет та, в которой находится <программа>.

TempMode=[Запрос][@Заголовок] 'Если в сценарии задана эта команда, то будет создана временная папка, в нее из архива извлекутся файлы, запущена программа, указанная в строке **Setup**, а после ее завершения временная папка со всеми файлами в ней будет удалена. Если задана команда **TempMode**, то в сценарии следует использовать команду **Setup** и нельзя применять команду **Path**.

В команде **TempMode** допустим необязательный строчный параметр, который будет для определения строк запроса и заголовка запроса перед началом извлечения. Если этот параметр присутствует, то извлечение будет начато только в том случае, если пользователь в ответ на запрос нажмет кнопку **Да**.

Пример полного сценария установки:

Title=Калькулятор 3.05 бета

Text=Калькулятор – условно-бесплатная программа.

Text=Вы можете использовать эту программу на протяжении

Text=пробного периода сроком в 40 дней.

Text=Если вы хотите продолжать пользоваться

Text= Калькулятором, то ДОЛЖНЫ зарегистрироваться

Text= в течение пробного периода.

Path=Calculator

Overwrite=1

Setup=setup.exe

TempMode=Продолжить установку Калькулятора

3.05?@Установка калькулятора

Непрерывный архив — это архив, в котором все сжимаемые

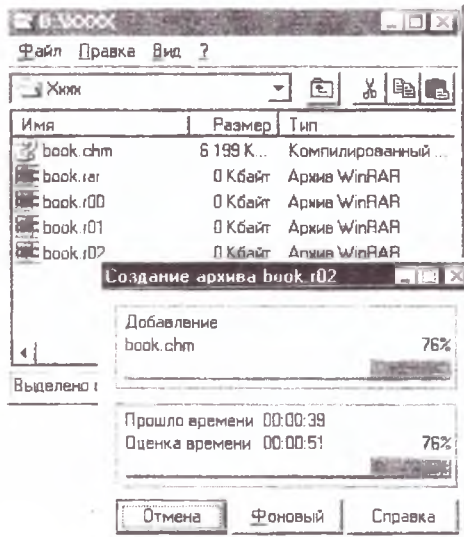


Рис. 8. Создание многотомных архивов

файлы рассматриваются как один последовательный поток данных. Они поддерживаются только в формате RAR. В диалоговом окне *Имя и параметры архива* устанавливается флаг *Создать непрерывный архив* (рис. 7).

Многотомное архивирование поддерживается только в формате RAR. Тома — это фрагменты архива, состоящего из нескольких частей.

В диалоговом окне *Имя и параметры архива* установите поле *Размер тома* (рис. 7). Первый том в последовательности имеет расширение *.rar*, последующие тома — *.r00*, *.r01*, *.r02* и т.д. (рис. 8). Тома могут быть также непрерывными и самораспаковывающимися. Файлы в существующих томах невозможно добавлять, обновлять или удалять.

2.3. Архиватор WinARJ

Для создания нового архива введите имя архива в поле *Archive Name*: (рис. 9). В поле *Destination Drive* укажите, куда будут сохраняться архивные файлы.

Чтобы добавить файлы к архиву, их необходимо выделить с использованием флага *sel.Extension* или флага *selected Files* (рис. 9). Флаг *Sel.Extension* позволяет выделять файлы через маски (**.exe*, **.**).

A *selected Files* позволяет выбирать отдельные файлы. После выделения файлов нажмите **Continue**.

При создании самораспаковывающихся архивов установите флаг *SelfExt Arch* (рис. 10).

Чтобы выполнить многотомное архивирование, установите флаг *Multi Volum* (рис. 10) и выберите размер тома. Предлагаются следующие размеры томов *360k*, *720k*, *1.2M*, *1.44M* или может быть установлен самостоятельно.

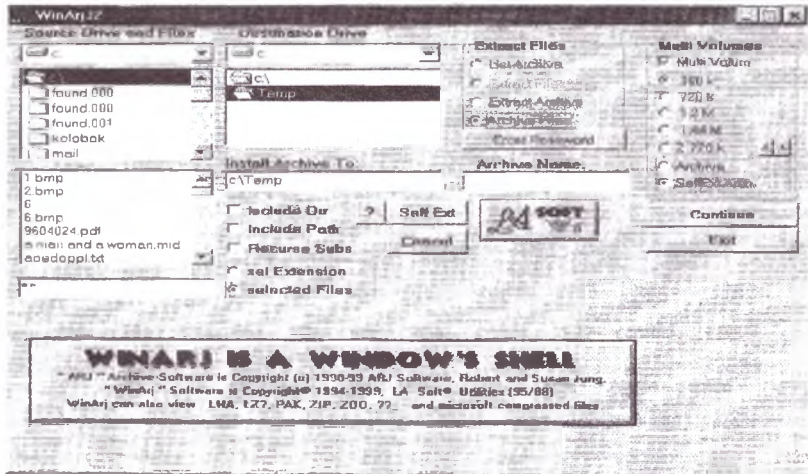


Рис. 9. Основные окна архиватора WinArj32

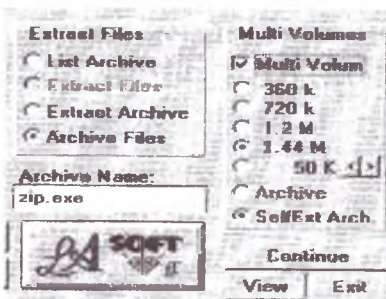


Рис. 10. Многотомное архивирование

2.4. Задание №2

1. Откройте программу WinZip

2. Создайте каталог с 5-ю файлами: текстовым, двумя графическими *jpg* или *gif*, документом Word (*.doc* или *.rtf*) и исполняемым (*.exe*, *.com*). В программе **WinZip** создайте новый архив *archiv.zip*. Добавьте в него все файлы текущей папки.

3. Через интерфейс *Моего Компьютера* или *Проводника* добавьте к архиву *archiv.zip* файлы с расширением *.doc* из папки */Мои документы*.

4. Добавьте к нему еще два любых файла через контекстное меню.

5. Выйдите из программы WinZip. Откройте созданный архив.

6. Преобразуйте архив *archiv.zip* в самораспаковывающийся.

7. Разбейте его на дискеты.

8. Откройте программу **WinRAR**.

9. Создайте непрерывный архив *x.rar*, включите в него все файлы папки */Мои документы*.

10. Преобразуйте *x.rar* в самораспаковывающийся архив.

11. Создайте самораспаковывающийся **GUI** архив *program.exe* на основе предложенной преподавателем инсталляции. В сценарии установки укажите текст, заголовок, путь распаковки, запрос на извлечение и программу установки.

12. Создайте в папке *C:/Temp* многотомный архив *DISKС* в программе **WinRAR**, содержащий все файлы, записанные на диске *C:* вне папок. Размер томов должен соответствовать *600Кб*.

13. Откройте архиватор **WinArij32**.

14. Создайте архив *document.arj* на диске *D:\Temp*. Добавить в него все файлы из папки */Мои документы*.

15. Создайте самораспаковывающийся архив *excel.exe* с размером тома *100Кб*, содержащий все файлы с расширением *.xls* из папки */Мои документы*.

3. ПРОСМОТР, ИЗВЛЕЧЕНИЕ, УДАЛЕНИЕ ФАЙЛОВ АРХИВА

3.1. Архиватор WinZip

Для просмотра и печати файлов архива используйте команду **Actions/View**, кнопку **View**, перетаскивание или открытие архива. После этого вызывается диалоговое окно *View*. Для печати списка файлов из архива выберите в меню **File/Print**.

Извлечение файлов из архива выполняется командой **Actions/Extract** или кнопкой **Extract** на панели инструментов. Файлы извлекаются в папку, указанную в поле *Extract to*. В диалоговом окне *Extract* имеются следующие возможности:

- *Selected files* – извлекать только те файлы, которые выбраны в основном окне.
- *All files* – извлекать из архива все файлы.
- *Files* – позволяют выбирать файлы по названию.
- *Overwrite existing files* – будут ли выдаваться подсказка до того, как файлы из архива будут записаны поверх существующих.
- *Skip older files* – будут ли не извлекаться существующие файлы, если дата/время создания файла в архиве более ранние, чем у файла на диске.
- *Use folder names* – будут ли использоваться хранящиеся в архиве названия папок.

Удаление файлов архива выполняется командой **Actions/Delete** или нажатием клавиши **Del**. *Эти операции нельзя отменить.*

Для управления архивом существуют команды **Copy Archive** (копирование открытого архива в другую папку или на другой диск), **Move Archive** (перенос архива), **Rename Archive** (изменить название архива) и **Delete Archive** (удалить архив) из меню **File**.

3.2. Архиватор WinRAR

Чтобы просмотреть файлы архива, выполните команду **Команды/Просмотреть файлы**, нажмите кнопку **Просмотр** на панели инструментов или комбинацию клавиш *Alt-V*. Команда доступна в режимах управления файлами и архивами.

Информация об архиве извлекается командой **Команды/Информация об архиве**, кнопкой **Инфо** или комбинацией клавиш *Alt-I*. Эта команда доступна только в режиме управления архивами.

Для извлечения файлов из архива выделите в архиве извле-

каемые файлы и выполните команду **Команды/Извлечь** в указанную папку, нажмите кнопку **Извлечь** в или комбинацию клавиш *Alt-E*. Команда доступна в режиме управления архивами.

Удаление файлов доступно как в режиме управления файлами, так и архивами. Для удаления выполните команду **Команды/Удалить файлы**, или нажмите кнопку на панели инструментов **Удалить**, клавиши *Alt-D* или *Del*.

3.3. Архиватор WinARJ

Для просмотра архива служит встроенный просмотрщик *VIEWER*, позволяющий просматривать текстовые файлы (из *.arj* архива) без выделения архива. В поле *Source-File List* необходимо выбрать архив (рис. 11). Выбираете архив, который необходимо просмотреть и **Continue**. Также можно послать файл на принтер или скопировать в буфер.

Filename	Original	Compressed	Ratio	Date	Time	modified	CRC-32	Attr	CPMGOX
ABOUT.FRM	7512	1661	0.221	94-12-17	20:35:30	CE9E46C9	---	W	B+1
ABOUT.FRX	11936	1442	0.121	94-12-17	20:35:26	F4894382	---	W	B+1
ARJ.PIF	545	142	0.261	94-12-08	09:53:36	6251F47D	---	W	B+1
ARJLIST.BAT	93	78	0.839	95-01-20	11:06:16	1825579C	A--	W	B+1
ARJMSC.FRM	6680	1802	0.270	94-12-21	18:50:16	36189A2E	---	W	B+1
ARJMSC.FRX	11166	1145	0.103	94-12-21	18:50:12	24C01303	---	W	B+1

Рис. 11. Просмотр архива

При извлечении файлов из архива выделите извлекаемые архивы, укажите директорий, куда будут извлекаться архивы в поле *Destination Drive*, установите флаг *Extract Archive*, нажмите **Continue** (рис. 9).

Для извлечения файлов из нескольких дисков выделите извлекаемые архивы, выберите директорий назначения, *Extract Files to:*, установите флаг *Multiple Disk* из поля *Extract Archives from* и нажмите **Continue**. Расширение файлов может быть только *.exe* или *.arj*.

3.4. Задание №3

1. Откройте архиватор *WinZip*.
2. Скопируйте архив *archiv.zip*, переименуйте его в *all.zip*.
3. Просмотрите все графические файлы различными способами.

Распечатайте один из текстовых файлов.

4. Создайте новую папку *ALL* и извлеките из *all.zip* исполняемые и *.doc* файлы (использовать хранящиеся в архиве названия папок).

5. Удалите из архива *.doc* файлы. Распечатайте список файлов архива на принтере.

6. Закройте программу **WinZip**, войдите в *Проводник*, дважды щелкните по *all.zip*, просмотрите содержимое.

7. Удалите архив *all.zip*.

8. Откройте архиватор **WinRAR**.

9. Просмотрите любой из *.doc* файлов в архиве *x.exe*.

10. Получите информацию об архиве *x.rar*. В отдельный текстовый файл *info.txt* запишите общее количество файлов, общую степень сжатия, общий размер несжатых файлов. Добавьте *info.txt* к *x.rar*.

11. Откройте архив *DISKC*, извлеките из него в папку *C:\Мои документы* файл *autoexec.bat*. Удалите архив *DISKC*.

12. Удалите файлы **.rtf* из архива *x.exe*.

13. Откройте архиватор **WinAri**.

14. Просмотрите содержимое файлов архива *document.arj*. Извлеките из него в папку *C:/Task* все файлы, размер которых не превышает 300Кб.

15. Запишите в свою папку на сетевом диске архив *document.arj*.

4. СПЕЦИАЛЬНЫЕ ВОЗМОЖНОСТИ АРХИВАТОРОВ

4.1. Архиватор WinZip

Для проверки целостности архива используйте команду **Actions/Test**. Результаты операции выводятся в диалоговом окне *View*.

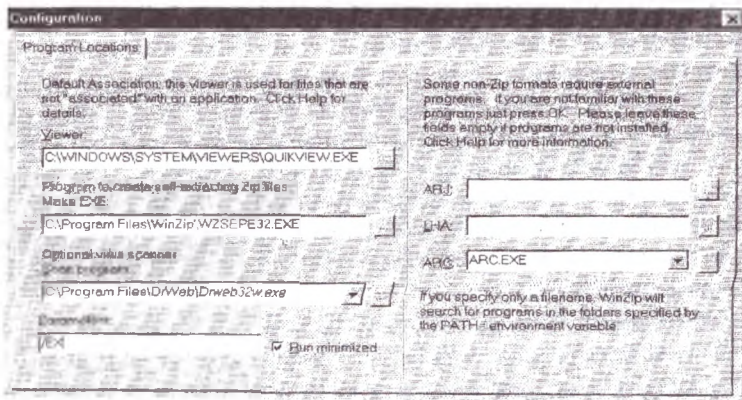


Рис. 12. Диалоговое окно Configurations

Команда **Actions/Virus Scan** проверяет архив на наличие вируса. При установке **WinZip** автоматически ищутся антивирусные программы (в системной переменной *PATH* и строках реестра Windows «*App Paths*»). Настройка антивирусной программы производится в диалоговом окне *Configuration | Program Locations* (рис. 12). Необходимо заполнить следующие поля: название антивирусной программы в поле *Scan program*; параметры запуска в поле *Parameters*; сворачивание окна в *Run minimized*.

Комментарии к архиву добавляются командой **Actions/Comment** через диалоговое окно *Comment*, позволяющее набрать, просмотреть, отредактировать и удалить текст комментария.

Защита zip-файлов паролем охраняет их от вмешательства случайных пользователей. Тем не менее, используемый формат шифрования не столь надежен и не может защитить от компрометации. В основном окне **WinZip** рядом с файлами, защищенными паролем, стоит знак плюс.

Чтобы защитить файлы паролем, необходимо указать его **ПОСЛЕ** открытия или создания архива, но **ДО** добавления файлов. В

диалоговом окне *Add* или *Drop* щелкните по кнопке **Password** и напечатайте пароль. Установить пароль также позволяет команда **Options/Password**.

Команда **Options/Configuration** активизирует диалоговое окно *Configuration*. В нем собрана большая часть настроек **WinZip**. Команда **Options/Sort by** позволяет определить порядок сортировки списка файлов. Команда **Options/Reuse WinZip windows** из выпадающего меню определяет, будет ли открываться новое окно **WinZip** при двойном щелчке по *zip*-файлу. Команда **Options/Save settings on exit** определяет, будут ли выбранные параметры автоматически сохраняться при закрытии программы. Команда **Options/Save settings now** моментально сохраняет выбранные параметры. Команда **Options/Set Installation Defaults** присваивает многим параметрам конфигурации и названиям папок значения по умолчанию. Команда **Options/View Last Output** служит для вывода на экран результатов выполнения последней операции **WinZip**.

Интерфейс WinZip Wizard. Мастер вызывается командой **File/Wizard** или кнопкой **Wizard** и дает следующие возможности: список *zip*-файлов из любимых папок (*Favorite Zip Folders*), упорядоченный по датам, автоматическая установка программ, распространяемых в виде *zip*-файлов, распаковка файлов в *Проводнике*, взаимодействие с *Netscape Navigator/Communicator* и *Microsoft Internet Explorer*.

Интерфейс Диспетчера файлов и Проводника с WinZip позволяет открывать архивы двойным щелчком в *Диспетчере файлов*, открывать **WinZip** через выпадающее меню *Диспетчера файлов* и автоматически открывать и закрывать **WinZip** вместе с *Диспетчером файлов*. Архивировать и распаковывать файлы можно в *Моем компьютере* и *Проводнике*.

Для архивирования **E-mail** служит команда **Zip and E-Mail filename.zip** контекстного меню файлов в *Моем компьютере* и *Проводнике*. Она позволяет сжать выбранные файлы и папки во временный *zip*-архив, создать e-mail сообщение и приложить к нему созданный архив. Временный архив затем удаляется автоматически.

Команда контекстного меню **Zip and E-Mail Plus** предоставляет те же функции, что и предыдущая и дополнительно позволяет задать имя и пароль *zip*-архиву.

Команда **File/Mail Archive** позволяет послать по e-mail открытый в текущий момент времени архив. При этом **WinZip** создает но-

вое сообщение и прикладывает к нему архив.

4.2. Архиватор WinRAR

Для проверки целостности архива выделите в архиве файлы и папки и вызовите **Команды/Протестировать файлы в архиве**, кнопку **Тест** на панели инструментов или комбинацию клавиш *Alt-T*. Найденные ошибки будут отображены в окне *Диагностические сообщения*.

Защита архивов от повреждений позволяет восстановить поврежденную информацию, используя специальный тип избыточной информации, называемый информацией для восстановления.

Необходимо включить опцию *Информация для восстановления* в диалоге ввода имени архива и его параметров (рис. 7) или включить аналогичную опцию при вызове команды **Создать по умолчанию** на вкладке *Архивация* из меню **Параметры/Установки** (рис. 13).

Для восстановления служит команда **Команды/Восстановить архив** или кнопка **Исправить** на панели инструментов. Восстанов-

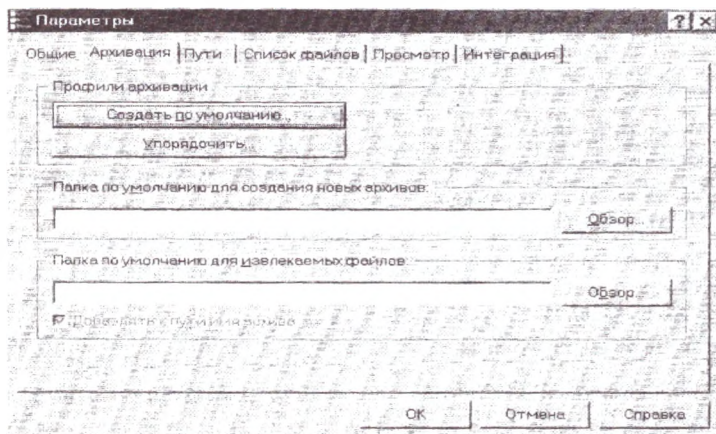


Рис. 13. Диалоговое окно Параметры

ленный архив получает имя *_recover.rar*. Если поврежденный архив не содержит информацию для восстановления или если его невозможно полностью восстановить, архиву присваивается имя *_reconst.rar*.

Команда **Команды/Добавить архивный комментарий**, кнопка

Коммент. или комбинация клавиш *Alt-M* добавляет **комментарии к архиву**.

Блокировка архивов выполняется командой **Команды/Заблокировать архив** в режиме управления архивами или комбинацией клавиш *Alt-L*. **Заблокированный архив невозможно изменить с помощью WinRAR.**

Чтобы **зашифровать** файлы, до архивации нужно указать пароль. Для ввода пароля – команда **Файл/Пароль** или щелкните на маленький значок ключа в левом нижнем углу окна **WinRAR**. Для удаления пароля введите пустую строку в диалоге ввода пароля или закройте **WinRAR** и снова его запустите.

Для **оценки степени сжатия** выполните **Команды/Оценить предполагаемую степень сжатия**, нажмите кнопку **Оценить** или комбинацию клавиш *Alt-S*. Приблизительно оценивается возможная степень и время сжатия файлов и папок с помощью разных форматов и методов сжатия.

Выбор оптимальных параметров архивации. При создании архива можно выбрать метод сжатия (рис. 7): *Без сжатия, Скоростной, Быстрый, Обычный, Хороший и Максимальный*, обеспечивающие различную степень и скорость сжатия.

Можно выбрать размер словаря. Большие значения параметра приводят к лучшему, но более медленному сжатию.

Мультимедиа-сжатие – дополнительный алгоритм для улучшения сжатия оцифрованного звука или графических изображений. Позволяет повысить степень сжатия на 30%.

Электронная подпись позволяет добавлять в архив электронную подпись с именем автора, временем последнего обновления и именем архива.

Интерфейс с Windows. **WinRAR** поддерживает применяемую в *Windows* технологию *Перетащить-и-оставить* (Drag & Drop).

4.3. Архиватор WinARJ

Защита паролем. Нажатие кнопки **Enter Password** позволяет зашифровывать и расшифровывать архивные файлы. В программе не предусмотрены пути восстановления утраченных паролей.

4.4. Задание №4

1. *Откройте архиватор WinZip.*
2. Проверьте архив *archive.exe* на целостность и наличие вируса.

3. Добавьте комментарий к архиву *archive.zip* об авторе и времени создания. Попробуйте добавить комментарий к архиву *archive.exe*.

4. Защитите *archive.zip* паролем и извлеките из него файлы.

5. Откройте архив *archive.exe* и отсортируйте файлы по коэффициенту сжатия. Обратите внимание, какие типы файлов имеют наибольший коэффициент.

6. Перейдите в **Wizard**, добавьте текущую папку в *Favorite Zip Folders*, удалите архив *archive.exe* и вернитесь в классический интерфейс. Закройте программу **WinZip**.

7. Через окно *Мой компьютер* добавьте один файл к архиву *archive.zip*. Просмотрите содержимое архива в *Проводнике*.

8. Создайте на сетевом диске свою папку, скопируйте туда *archive.zip*.

9. Откройте архиватор **WinRar**.

10. Скопируйте архив *x.rar* на другой диск под именем *doc.rar*.

11. Включите в архив *doc.rar* информацию для восстановления. Сравните размер архива с первоначальным.

12. Откройте *doc.rar* в текстовом редакторе и измените 25-й бит. Сохраните сделанные изменения. Восстановите архив.

13. Добавьте комментарий о содержимом и заблокируйте архив *x.rar*. Защитите паролем файл *info.txt*.

14. Оцените степень сжатия папки */Program files* и выберите оптимальные параметры архивации по времени и занимаемому размеру.

15. Выбрать один из графических файлов, оценить степень его сжатия, заархивировать файл, используя метод *Лучший (Быстрый)* и размер словаря *128Кб*. При сжатии установить флаг *Мультимедиа сжатие*.

16. Удалите архивы *doc.rar* и *program.exe*.

17. Сравнить размеры архивов *archive.zip*, *x.rar*, *document.arj*. Рассчитать для каждого процент сжатия.

5. WINDOWS BACKUP

В **Windows 2000 Backup 5.0.** существует возможность архивирования файлов. Такая функция обладает всеми свойствами настоящего приложения архивации, включая ежедневную архивацию по расписанию, добавочную и разностную архивацию.

5.1. Конфигурация задач резервного копирования

Для запуска **Windows Backup** выберите *Пуск/Программы/Стандартные/Служебные/Архивация данных.*

Выбранные настройки сохраняются в файле для создания дополнительных заданий по архивации. Выберите **Задание/Сохранить выбранные как (Jobs/Save Selection As)** и сохраните файл в папке, доступной всем пользователям системы (например, *Documents and Settings, All Users*).

Выберите **Сервис/Параметры (Tools/Options)** и перейдите на вкладку *Общие (General)*, где выполняются настройки под конкретную среду. Уберите флажок *Архивировать содержимое подключенных дисков (Back up the contents of mounted drives)*, чтобы избежать избыточного трафика. Перейдите на остальные вкладки и убедитесь, что конфигурация режима **Backup** соответствует задаче. На вкладке *Исключение файлов (Exclude Files)* приведен список файлов, которые программа игнорирует по умолчанию во время процедуры архивации. Для большинства систем этот список не требует изменений, но при необходимости можно добавить в него файлы или удалить. После того, как выбраны параметры архивации, укажите имя и местонахождение файла архива. Просмотрите папки и выберите подходящее место (рис. 14), или введите путь вручную. Перейдите на вкладку *Архивация (Backup)*, щелкните *Архивировать (Start Backup)* и в открывшемся окне *Сведения о задании архивация (Backup Job Information)* введите имя задания для архивации.

Поскольку перезапись файла с резервной копией является деструктивной процедурой, лучше использовать дополнительные меры защиты. Например, можно воспользоваться функцией *Добавить этот архив (Append this backup to the media)* для того, чтобы создать файл без перезаписи предыдущей версии, а затем использовать команду **Удалить (Delete)** после выполнения регулярной архивации сервера на ленту. Для настройки дополнительных параметров в окне *Сведения о задании архивации* щелкните **Дополнительно (Advanced)**. Выберите

— снова загрузятся сделанные ранее настройки. Проверьте конфигурацию, убедитесь, что выбраны правильный тип архивации и параметры *Append/Replace* и верно задано место хранения архива.

Далее, щелкните *Пуск/Программы/Стандартные/Службные/Назначенные задания (Start/Programs/Accessories/System Tools/Scheduled Tasks)* и убедитесь, что операции заданы корректно и что они работают. Запланированные задания работают в фоновом режиме независимо от того, выполнена регистрация в системе или нет.

5.3. Создание ярлыков для задач архивации

Ярлыки позволяют пользователям выполнять заранее настроенные задания по резервированию. В окне *Расписание (Scheduled Tasks)* (см. рис. 15) укажите задачу и выберите пункт **Свойства**

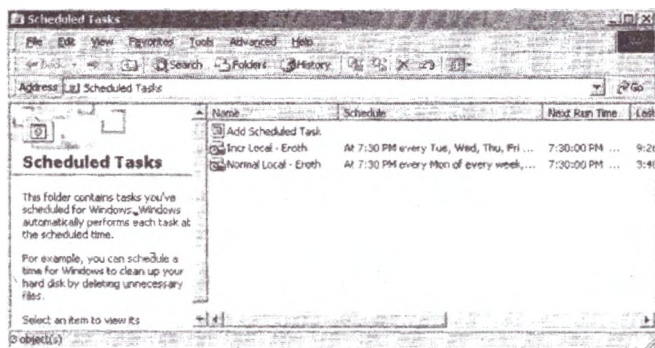


Рис.15. Задания планировщика

(**Properties**) для просмотра деталей настройки, затем скопируйте команду и аргументы из поля *Запуск (Run)* вкладки *Задание (Task)* в ярлык, который доступен пользователю.

Возможно, подогнать строку удастся, если переместить часть пути в поле *Start in*. Если это не поможет, скопируйте целую строку в текстовый редактор, например *Блокнот (Notepad)*, сохраните файл с расширением *.cmd* или *.bat* и создайте ярлык этому файлу.

Создав ярлык и сделав его доступным, убедитесь, что пользователь, который будет выполнять задачу, обладает соответствующими полномочиями. Самый простой способ назначить полномочия состоит в том, чтобы сделать учетную запись пользователя членом локальной

группы *Backup Operators*. В отличие от процедур архивации, которые выполняются через планировщика, эти ручные процедуры архивации будут видны пользователю и будут проходить под его контролем.

5.4. Извлечение файлов из архива и восстановление системы

Операция извлечения файлов предназначена для возврата потерянного или поврежденного содержимого, а операция восстановления – для ситуаций катастрофических сбоев, когда требуется переустановка системы.

Прежде чем можно будет восстановить что-либо, необходимо найти соответствующий файл и каталог. Для этого запустите **Backup**, перейдите на вкладку *Восстановление (Restore)*, щелкните *Файл (File)* и выберите *Catalog file*. Введите путь архивного файла и нажмите **ОК**. Затем выберите файлы для извлечения и установите параметры, которых должна придерживаться система во время извлечения. По умолчанию извлеченные файлы помещаются в их исходное местоположение.

Параметр *Альтернативное размещение (Alternate location)* позволяет просматривать содержимое файла архива пользователя со своего настольного компьютера и извлекать его содержимое, помещая на систему этого пользователя. Указанный путь является относительным, а структура папок восстанавливаемого фрагмента останется неизменной.

Если с помощью **Backup** производится полное восстановление системы, следует выполнить новую установку операционной системы и восстановить все содержимое архива в установленную операционную систему.

5.5. Задание №5

1. Запустите **Windows Backup**. Создайте и сохраните задание архивации «Резервирование документов», сохраняющее на сетевой диск в файл *Docbackup.bkf* папку */Мои документы*. Установите параметры задания и создайте отчет архивирования.

2. Настройте расписание для ежедневного вечернего запуска (если компьютер не используется) вашей задачи архивации.

3. Создайте ярлык для настроенного задания и сделайте его доступным всем пользователям.

4. Выполните извлечение файлов из архива *Docbackup.bkf* в папку *D:\Temp*.

ЛИТЕРАТУРА

1. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: ДИАЛОГ-МИФИ – 2002. – 384 с.
2. Морозов Н.П., Чернокнижный С.Б. Защита деловой информации для всех. – М.: Весь – 2003. – 159 с.
3. Зелинский С. В цифровых тисках // Chip. 2001. – № 9.
4. <http://www.compression.ru>
5. Шафрин Ю. Основы компьютерной технологии. – М.: АБФ. – 1996.
6. Кендл Т. Форматы файлов Internet. – СПб: Питер. – 1997.
7. <http://www.book.itep.ru/2/26/>
8. Техническая документация к архиваторам. User's Guide.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1. АЛГОРИТМЫ СЖАТИЯ ТЕКСТОВОЙ ИНФОРМАЦИИ	4
1.1. Код Шеннона – Фано	4
1.2. Код Хаффмана	4
1.3. Арифметические коды	6
1.4. Словарные алгоритмы	7
1.5. LZ алгоритм	7
1.6. Задание №1	10
2. СОЗДАНИЕ АРХИВОВ	11
2.1. Архиватор WinZip	11
2.2. Архиватор WinRAR	12
2.3. Архиватор WinARJ	15
2.4. Задание №2	16
3. ПРОСМОТР, ИЗВЛЕЧЕНИЕ, УДАЛЕНИЕ ФАЙЛОВ АРХИВА	18
3.1. Архиватор WinZip	18
3.2. Архиватор WinRAR	18
3.3. Архиватор WinARJ	19
3.4. Задание №3	19
4. СПЕЦИАЛЬНЫЕ ВОЗМОЖНОСТИ АРХИВАТОРОВ	21
4.1. Архиватор WinZip	21
4.2. Архиватор WinRAR	23
4.3. Архиватор WinARJ	24
4.4. Задание №4	24
5. WINDOWS BACKUP	26
5.1. Конфигурация задач резервного копирования	26
5.2. Настройка планировщика архивации	27
5.3. Создание ярлыков для задач архивации	28
5.4. Извлечение файлов из архива и восстановление системы	29
5.5. Задание №5	29
ЛИТЕРАТУРА	30

Учебное издание

СЖАТИЕ И АРХИВИРОВАНИЕ ДАННЫХ

Учебно-методическое пособие

Составители: **Пацей** Наталья Владимировна
Урбанович Павел Павлович

Редактор К.В. Мишина

Подписано в печать 18.06.2004. Формат 60x84 1/16
Бумага офсетная. Гарнитура Таймс. Печать офсетная.

Усл. печ. л. 2,2. Уч.-изд. л. 1,9.

Тираж 100 экз. Заказ 392.

Учреждение образования «Белорусский государственный
технологический университет».
220050. Минск, Свердлова, 13а. Лицензия ЛВ № 276 от 15.04.2003.

Отпечатано в лаборатории полиграфии учреждения образования
«Белорусский государственный технологический университет».
220050. Минск, Свердлова, 13.