

Архивация файлов

Файлы и файловая
система

Избыточность

- Редакторы, работающие с текстовой, графической, звуковой и другой информацией, **кодируют** ее наиболее **естественным**, но **не самым экономичным** способом.
- Действительно, если внимательно посмотреть любой текст, то можно заметить, что такие буквы «а» и «о», встречаются в нем гораздо чаще чем «ю» и «у». То же самое можно отнести и к сочетаниям букв.
- На рисунках цвета соседних точек в большинстве случаев близки по оттенку. Подобно этому в любой последовательности информации некоторые сочетания встречаются намного чаще других.

Избыточность

- Все это приводит к тому, что в файлах, хранящих эту информацию, некоторые **комбинации** из 0 и 1 **встречаются гораздо чаще**, чем другие. В таких случаях говорят, что информация обладает избыточностью, и есть возможность перекодировать содержание файла, уменьшив его размер.
- Для **сжатия** достаточно придерживаться правила: чем чаще встречается комбинация, тем более коротким сочетанием из 0 и 1 ее можно **перекодировать**. Разумеется, делать это должна программа.

Архивация и сжатие файлов

Архивация – создание резервных копий (на CD, DVD). Цели:

- сохранить данные на случай сбоя на диске
- объединить группу файлов в один архив
- зашифровать данные с паролем

Сжатие файлов – это уменьшение их размера. Цели:

- уменьшить место, которое занимают файлы на диске
- уменьшить объем данных для передачи через Интернет

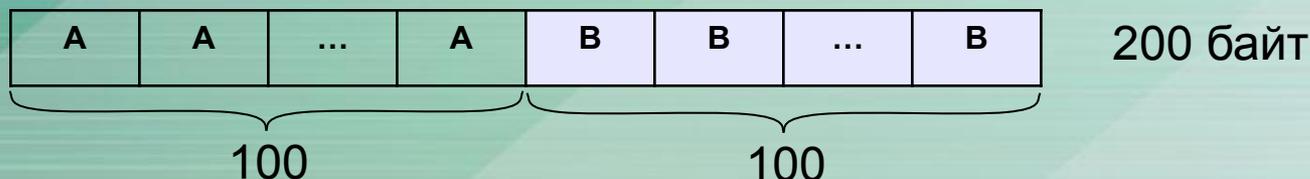
Типы сжатия:

- **без потерь:** сжатый файл можно восстановить в исходном виде, зная алгоритм сжатия
 - тексты
 - программы
 - данные
- **с потерями:** при сжатии часть информации безвозвратно теряется
 - фотографии (* .jpg)
 - звук (* .mp3)
 - видео (* .mpg)

Почему файлы можно сжать?

Алгоритм RLE (англ. *Run Length Encoding*, кодирование цепочек одинаковых символов, используется для рисунков *.bmp)

Файл qq.txt



Файл qq.rle (сжатый)



сжатие в 50 раз!



Сжатие с потерями или без?

Сжатие возможно, если в данных есть повторяющиеся символы или цепочки символов, сжатие «устраняет» эту **избыточность**.

Почему файлы можно сжать?

Общий подход:

- найти в данных повторяющиеся цепочки символов
- обозначить их короткими кодами (битовыми, разной длины)
- в начало сжатого файла записать словарь

Эффективные алгоритмы:

- алгоритм Хаффмана
- алгоритм LZW (Лемпела-Зива-Велча)
- алгоритм PPM (WinRAR)

Сжимаются

хорошо

- тексты (*.txt)
- документы (*.doc, *.xls)
- несжатые рисунки (*.bmp)
- несжатый звук (*.wav)
- несжатое видео (*.avi)

плохо

- случайные данные
- программы (*.exe)
- архивы (*.zip, *.rar)
- сжатые рисунки (*.gif, *.jpg, *.png, *.tif, ...)
- сжатый звук (*.mp3, *.wma)
- сжатое видео (*.mpg, *.wmv)

Специальные типы архивов

SFX-архив (англ. *Self eXtracting* – самораспаковывающийся) – это файл с расширением ***.exe**, который содержит сжатые данные и программу распаковки (около 15 Кб).



- для распаковки не нужен архиватор
- может распаковать неквалифицированный пользователь



- увеличение размера файла
- опасность заражения вирусами

Многотомный архив – это архив, разбитый на несколько частей. **Цели:**

- перенос через дискеты
- удобство скачивания через Интернет

WinRAR:

- `abc.part1.rar`, `abc.part2.rar`,
- многотомный SFX-архив: `abc.part1.exe`, `abc.part2.rar`,

Архиватор WinRAR (Е. Рошал)

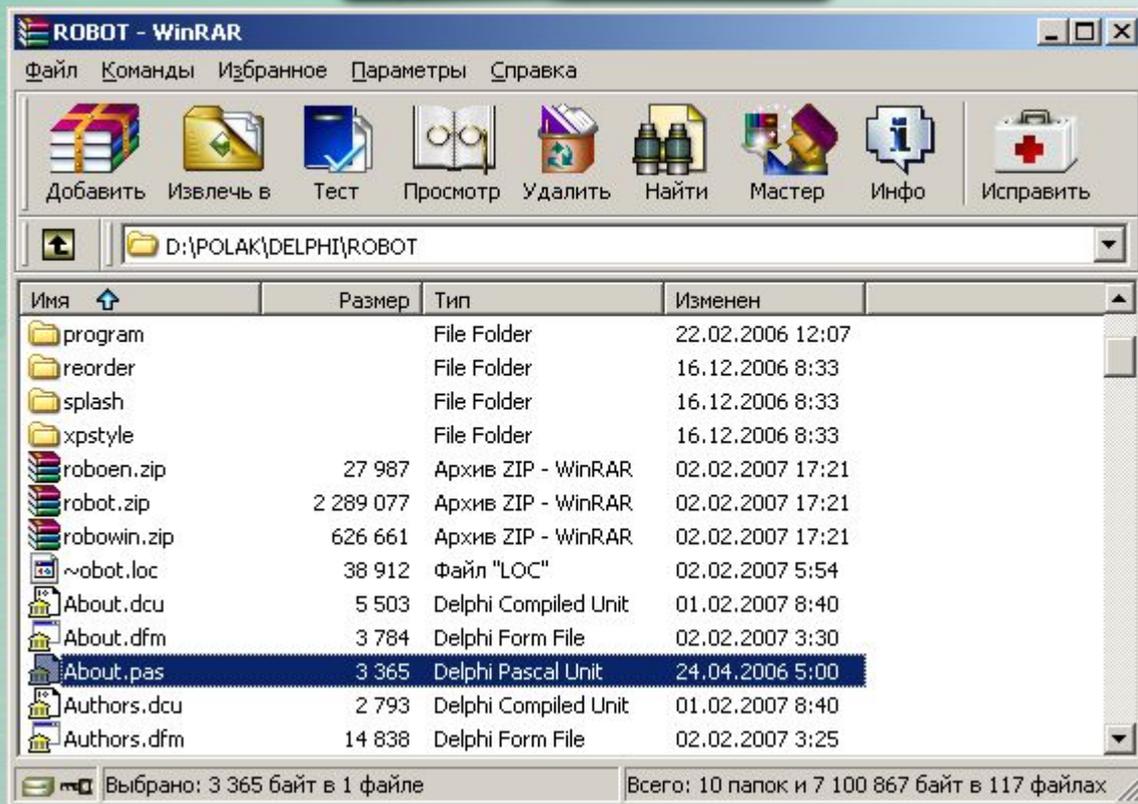
Запуск: Пуск – WinRAR

распаковать архив

сжать выделенные файлы

выйти из папки

двойной щелчок ЛКМ: войти в архив

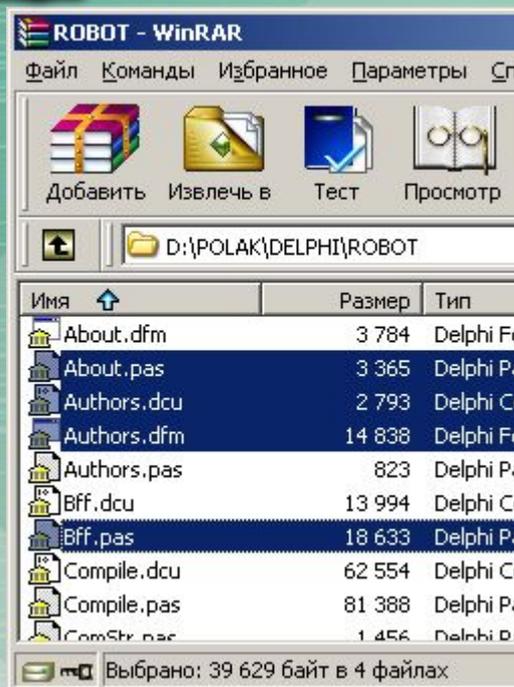


сменить диск

изменить пароль

Архиватор WinRAR: упаковка

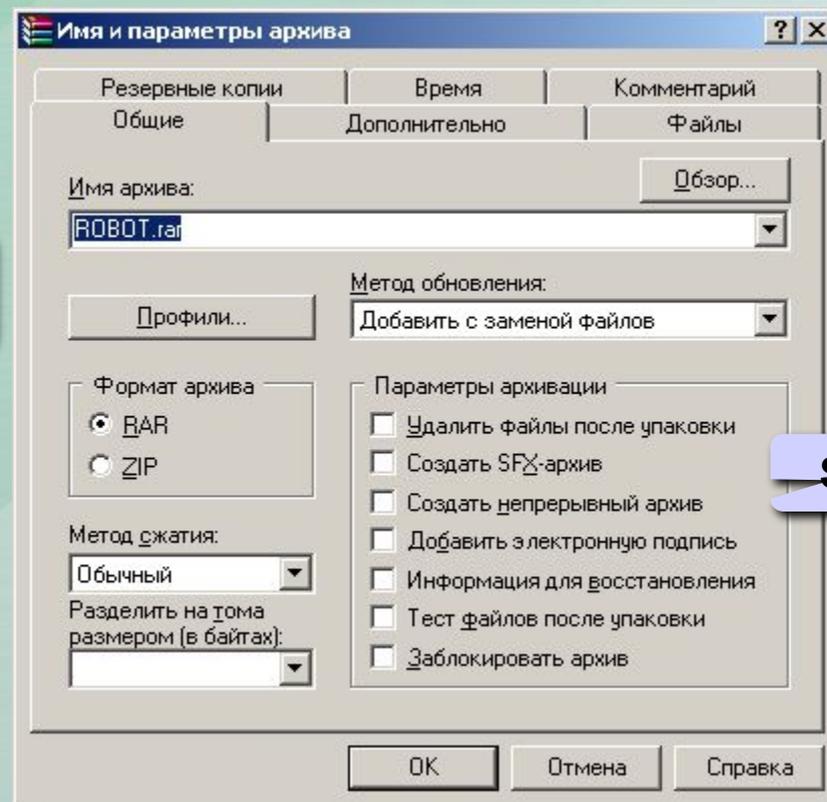
ЛКМ



Имя
архива

пароль

тип
архива

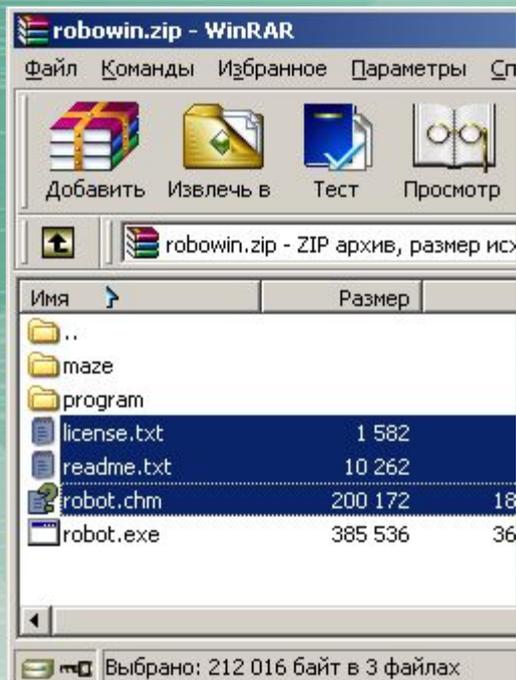


SFX

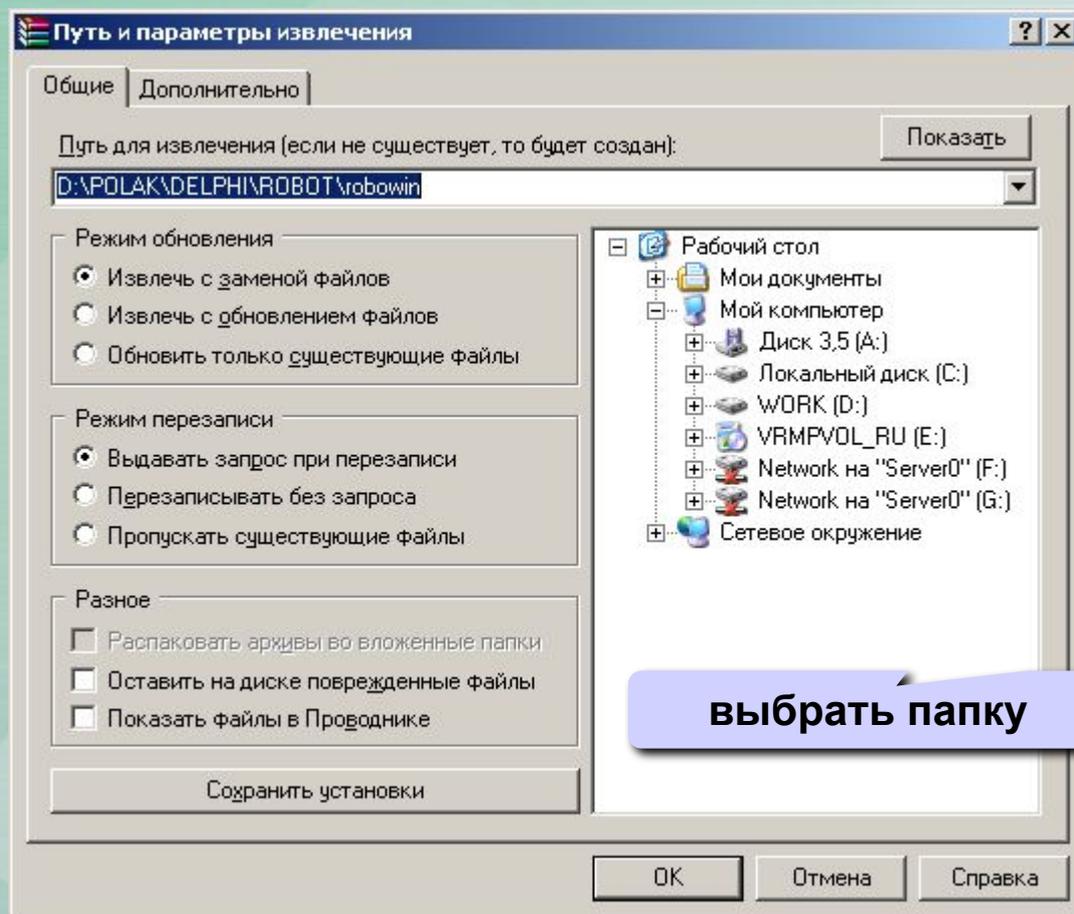
МНОГОТОМНЫЕ
архивы

Архиватор WinRAR: распаковка

ЛКМ



куда распаковать?

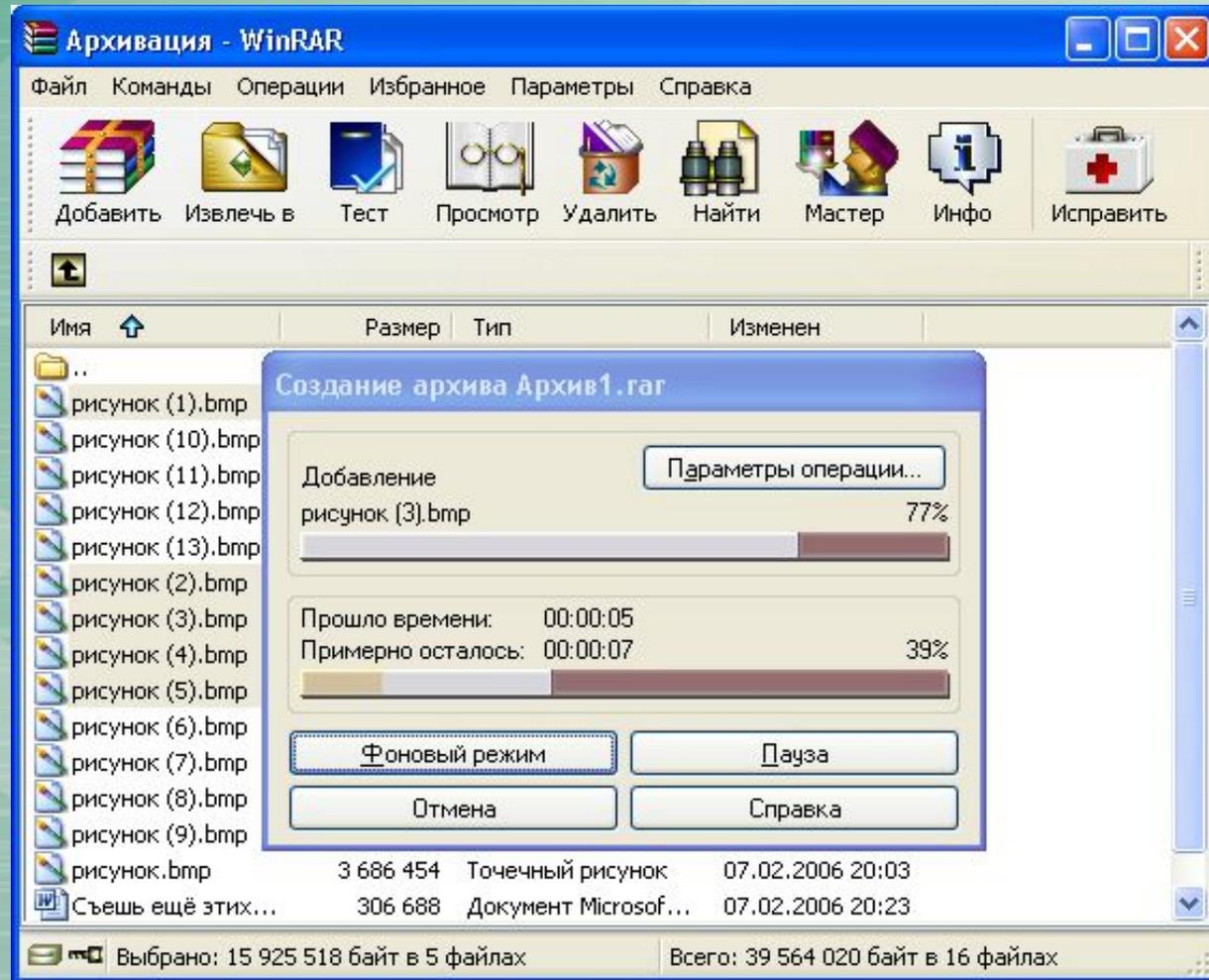


выбрать папку

Основные действия при работе с архивами:

- создание нового архива;
- добавление файлов в архив;
- просмотр содержимого архива;
- извлечение файлов из архива;
- просмотр файла в архиве;
- удаление файлов из архива.

Архивация с помощью оболочки WinRAR



Информация об архиве

Архив Архив1.rar

Информация | Параметры | Комментарий | SFX

RAR архив

Версия для извлечения: 2.9
Базовая ОС: Windows

Всего файлов: 5
Общий размер: 15 925 518
Размер в архиве: 5 715 310
Степень сжатия: 35%

Размер SFX-модуля: 0 байт
Главный комментарий: Нет
Пароли: Нет

Размер словаря: 4096 Кб
Информация для восстановления: Нет
Блокировка архива от изменений: Нет

Электронная подпись: Нет

35%

OK Отмена Справка

Архив Съешь ещё этих мягких французск...

Информация | Параметры | Комментарий | SFX

RAR архив

Версия для извлечения: 2.9
Базовая ОС: Windows

Всего файлов: 1
Общий размер: 306 688
Размер в архиве: 4 276
Степень сжатия: 1%

Размер SFX-модуля: 0 байт
Главный комментарий: Нет
Пароли: Нет

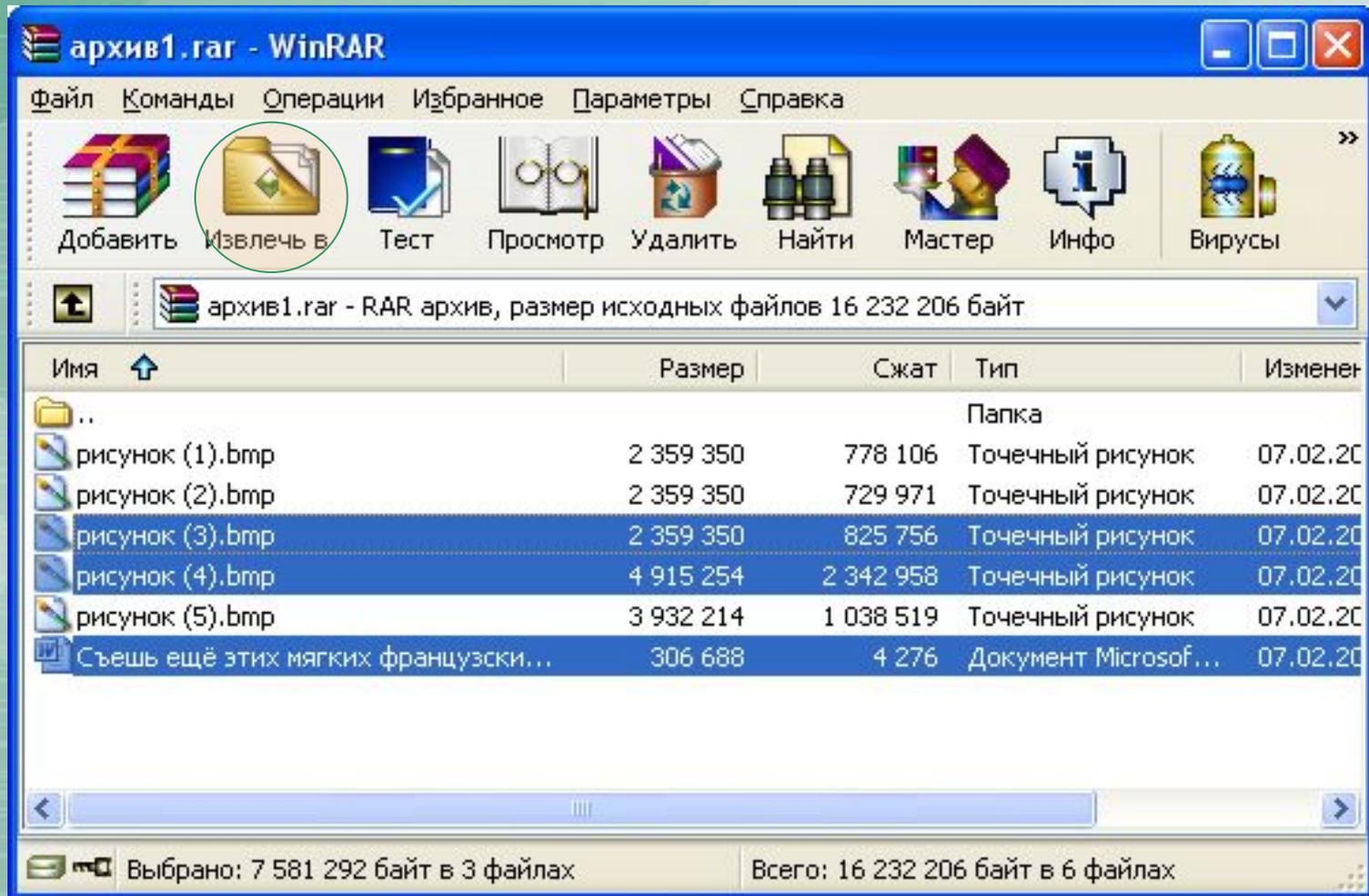
Размер словаря: 512 Кб
Информация для восстановления: Нет
Блокировка архива от изменений: Нет

Электронная подпись: Нет

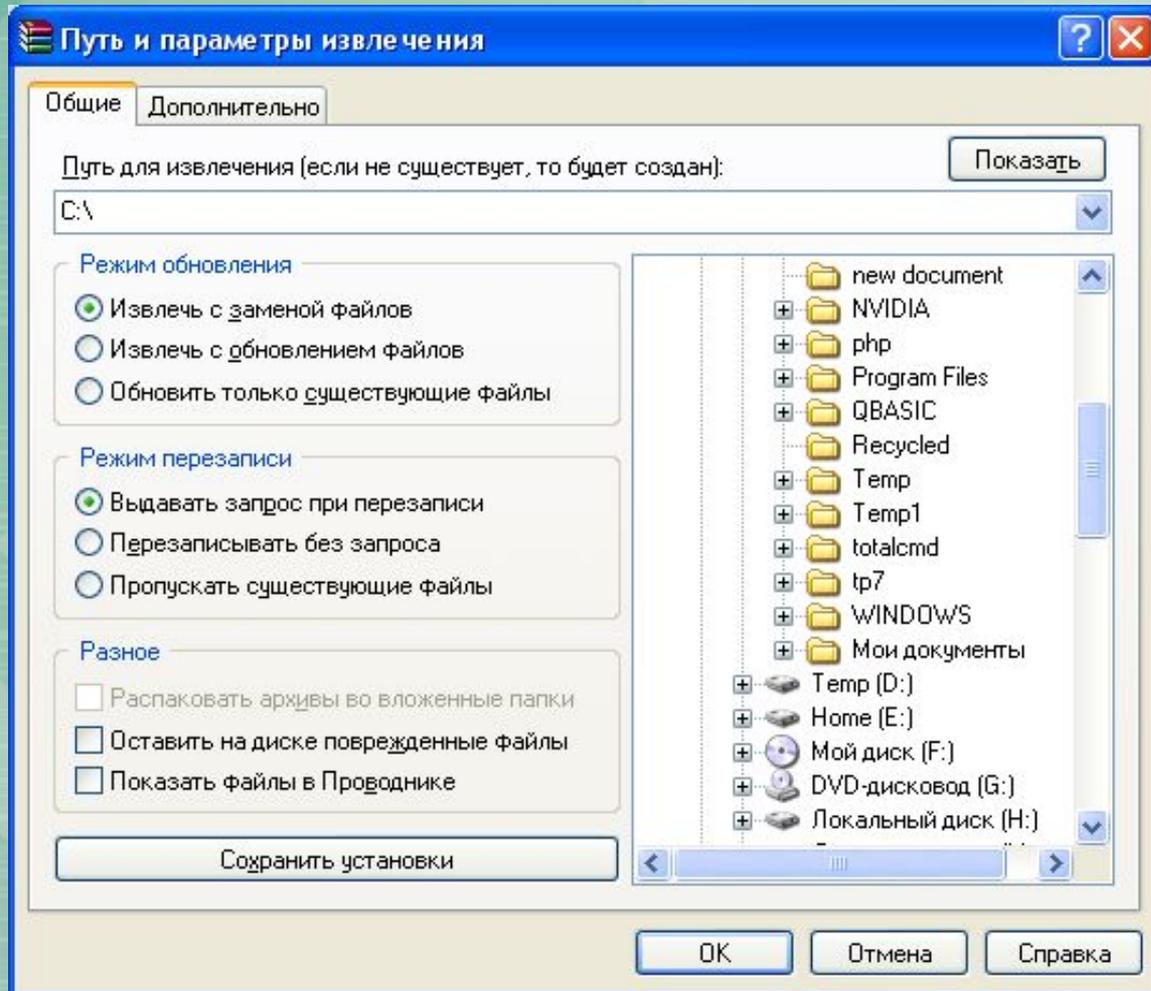
1%

OK Отмена Справка

Распаковка файлов



Распаковка файлов



Вопросы:

- Почему есть возможность уменьшать размер файлов?
- Что такое архивация?
- Какие файлы не имеет смысла архивировать?
- Почему перед пересылкой текстового файла по электронной почте имеет смысл предварительно его упаковать в архив?