

III. Архивиране на данни

1. Архивиране и компресиране на данни

Поради активната работа с компютри в наши дни, се забелязва непрекъснато нарастване на броя на използваните файлове и техния обем. Да се справим по-лесно с големия брой и големите размери на файловете ни помагат технологиите, наречени *архивиране* (англ. archiving) и *компресиране* (англ. compressing).

Архивиране

Да се *архивира* група от файлове означава да се създаде **нов файл**, който съдържа копие на всеки един от архивираните файлове и който се нарича *архив* (англ. archive). С архивираните файлове не можем да си служим директно, но когато се налага, всеки от тях може да бъде извлечен от архива, в обичайния му формат, и да бъде използван според предназначението му. Обратното действие – възстановяването на оригиналния файл от копието в архива – наричаме *разархивиране*.

Архивирането на данни се налага, най-вече, за да не се претрупва твърдият диск на компютъра с **документи и програми, които се използват сравнително рядко**. Обичайно е, такива програми и документи да се архивират и да се съхраняват на външен носител (CD, DVD).

Архивиране си струва да използваме и когато разполагаме с **група от файлове с общо предназначение** и при работа с тях, а най-вече при пренасянето им от едно място на друго, бихме искали всички тези файлове да са налични. Да си припомним как при създаване на компютърна презентация с обемисти видео-материали, програмата MS PowerPoint не вмъква тези материали в презентацията, а ги **свързва** с нея. Когато искаме да пренесем презентацията на друг компютър, трябва заедно с *.ppt файла да пренесем и всички свързани с него файлове. В такъв случай, само събирането на необходимите ни файлове в архив ще гарантира успешното пренасяне на презентацията.

За работа с архиви са необходими специализирани програми, които се наричат *архивиращи програми* (или *архиватори*). С такава програма може да се архивира един файл, група от файлове, една или няколко папки и дори цял твърд диск. Архивирането на твърд диск се използва по-рядко.

Основните функции на всяка архивираща програма са:

- *създаване* на архив с посочени от потребителя файлове;
- *добавяне* на нови файлове в архива;
- *подмяна* на архивирани файлове с техни нови версии;
- *изтриване* на файлове от архива;
- *справка* за съдържанието на архива и др.

Различните архивиращи програми имат и свои специфични функции. С възможностите на архивиращата програма 7-zip, която използвахме в урока за инсталиране и деинсталиране, ще се запознаем в този и в следващия урок.

Компресиране на данни

Компресиране на данни наричаме такова представяне на данните, че при съхраняването им върху носител те да заемат по-малко място от обичайното. Могат да се посочат много примери, когато са налага да се компресират данни. Например, когато трябва да копираме данни на външен носител, чийто обем е по-малък от размера на данните или когато искаме да освободим част от дисковото пространство на твърдия диск.

За да се постигне необходимото намаляване на обема на данните, се използват различни алгоритми за компресия, някои от които са създадени на базата на много сериозен математически апарат. Ще дадем прост пример на алгоритъм за компресия.

Всеки байт може да съдържа цяло число в интервала от 0 до 255. Даден е файл, съдържащ поредицата от 18 байта:

2, 2, 2, 4, 4, 4, 4, 4, 4, 4, 33, 14, 25, 21, 21, 21, 21, 21.

Ще използваме следния прост **алгоритъм за компресиране**. Тъй като файлът започва с три еднакви байта, съдържащи числото 2, ще поставим в началото на компресирания файл числата 3 и 2 (означаващи 3 пъти байт със съдържание 2). Следващите 7 байта, съдържащи числото 4, ще компресираме с двойката числа 7, 4. Следващите три байта съдържат три различни числа – 33, 14 и 25. За такъв участък от файла ще поставяме в компресирания файл числото 0, последвано от броя на различните числа – 3, както и самите числа: 0, 3, 33, 14, 25. Поредицата от 5 еднакви байта в края на файла, съдържащи числото 21, компресираме до 5, 21. Така получаваме компресирания файл:

3, 2, 7, 4, 0, 3, 33, 14, 25, 5, 21,

съдържащ 11 байта. Тъй като $11:18 \approx 0,6$ казваме, че файлът в този случай е компресиран до 60% от първоначалния му обем или че е постигната **степен на компресия** 60%.

Алгоритъмът за декомпресиране е очевиден. Първото число в компресирания файл е 3. Значи следващият байт, със стойност 2, ще се повтори 3 пъти в декомпресирания файл: 2, 2, 2. Следва числото 7, което значи, че следващият байт 4 ще трябва да бъде поставен в декомпресирания файл 7 пъти: 4, 4, 4, 4, 4, 4, 4. Числото 0, последвано от числото 3, означава, че следващите 3 байта ще са с различни съдържания и трябва да пренесем тези три байта – 33, 14 и 25 – в декомпресирания файл. Байтът, съдържащ числото 5, означава, че трябва 5 пъти да поставим в декомпресирания файл следващото число – 21. В резултат на декомпресирането получаваме началния файл:

2, 2, 2, 4, 4, 4, 4, 4, 4, 4, 33, 14, 25, 21, 21, 21, 21, 21.

Не е трудно да се забележи, че ефектът на намаляване на обема на файла, при този алгоритъм, се постига заради наличието на **последователности от еднакви байтове**. Такива последователности се срещат най-често при съхраняване на **растерни графични изображения** в различните файлови формати, които познаваме.

Демонстрираната техника на компресиране можем да наречем *компресиране без загуби* (англ. lossless), защото в резултат на декомпресирането получаваме точно файла, който е бил подложен на компресиране. Когато се компресират графични изображения, видео и звукозаписи, може да се използват и по-сложни алгоритми, които за да постигнат по-голяма бързина и по-висока степен на компресия, извършват *компресиране със загуби* (англ. with losses). Малките отклонения от първоначалния файл, които съответните алгоритми допускат при такъв вид компресиране не са фатални, тъй като човешкото око и ухо не са в състояние да доловят подобни малки изменения в изображението или звука.

Ако трябваше да използваме посочения в алгоритъм за кодиране на растерно изображение, спокойно можеше да пренебрегнем наличието на трите различни байта в средата на файла и да заместим част от тях (например 2) с байта отляво, а останалите (в случая 1) – с байта отдясно. Така щяхме да получим компресирания файл:

3, 2, 9, 4, 6, 21,

който е с 30% от първоначалния обем или е постигната степен на компресия 30%.

Архивиране с компресия

Много програми използват компресиране на данните, с които работят. На практика повечето от графичните, звукови и видеоредактори компресират съответните обекти с повече или по-малко загуби. От много отдавна, обаче, **компресирането на данни е станало неотменна част от процеса на архивиране на данни**. Днес на практика няма архивираща програма, която да не използва компресиране. И това е напълно обяснимо, защото обикновено един архив съдържа много файлове и е с много голям обем. При това положение, ненужното на пръв поглед архивиране на един отделен файл добива смисъл – архивирането на такъв файл е равносилно на компресиране на файла.

Необходимостта от архивиране и компресиране на данни е осъзната отдавна от производителите и потребителите на софтуер и се предлага още при ОС с буквено-цифров (команден) интерфейс. В ОС Unix и подобните на нея (различните версии на Linux, например) са популярни архивиращите програми TAR и GZIP. Най-популярни архивиращи програми, които работеха в ОС DOS, бяха ARJ, AIN, RAR, PKZIP и PKUNZIP. За работа в ОС с графичен интерфейс най-използвани са WinZIP, WinRAR, MacZIP, SmartZIP и др. Използването на изброените програми изисква лиценз.

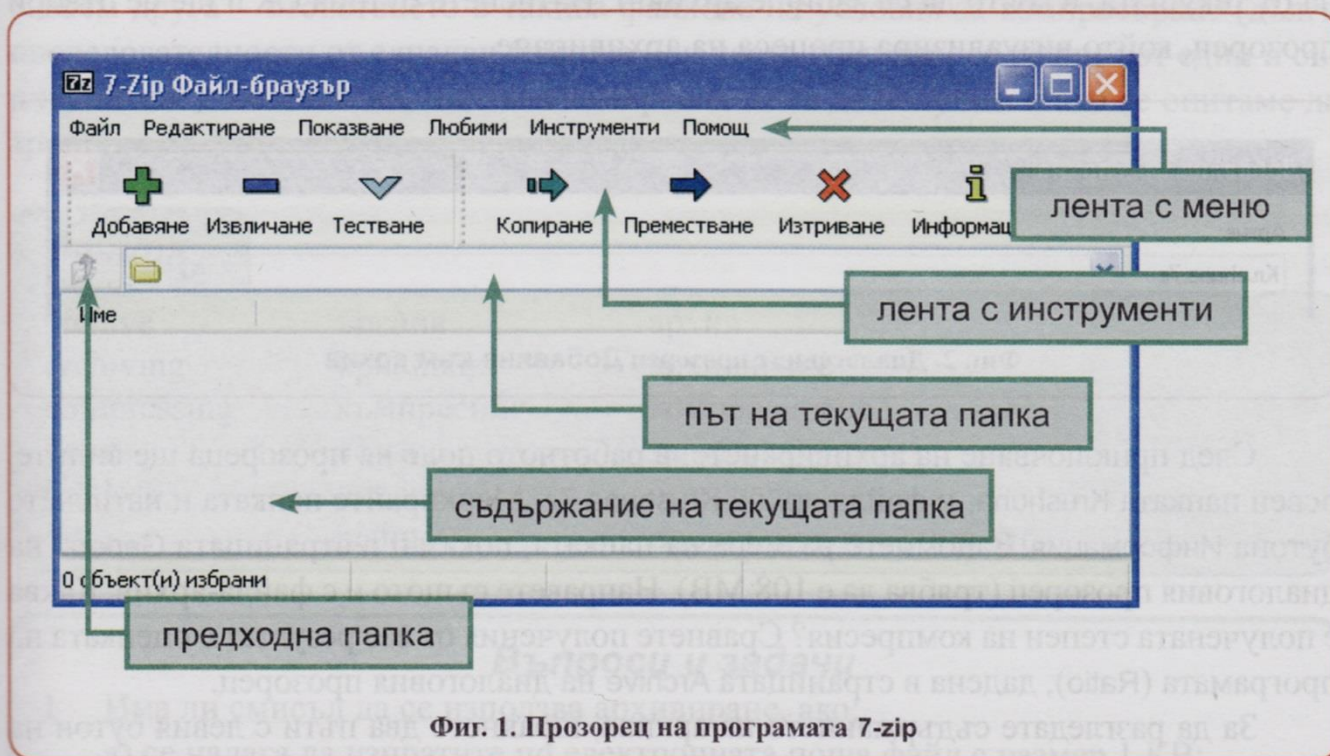
Когато архивът е направен с WinZIP, казваме че този файл е в ZIP-формат, когато е с WinRAR – че е в RAR-формат. Това са двата най-популярни архиватора и всеки от тях работи с архивите (форматите) на другия.

Архиваторът 7-zip

Ще се запознаем по-подробно с архиватора 7-zip, който инсталирахме на компютрите в предишен урок. Както вече знаем, тази програма се разпространява като свободен софтуер и има редица предимства:

- лесно се инсталира;
- поддържа различни формати за архивиране и компресия;
- интерфейстът с потребителя е подобен на този в Windows Explorer;
- за форматите ZIP и GZIP архиваторът 7-zip постига степен на компресия, която е до 10% по-добра от тази на PKZIP и WinZIP;
- предлага диалог на български език.

След инсталирането, архиваторът се стартира от бутона Start/All Programs/7-zip/7-zip File Manager. На *Фиг. 1* е показан прозорецът на програмата. Най-отгоре е лентата със стандартните за ОС Windows менюта – Файл (File), Редактира-



Фиг. 1. Прозорец на програмата 7-zip

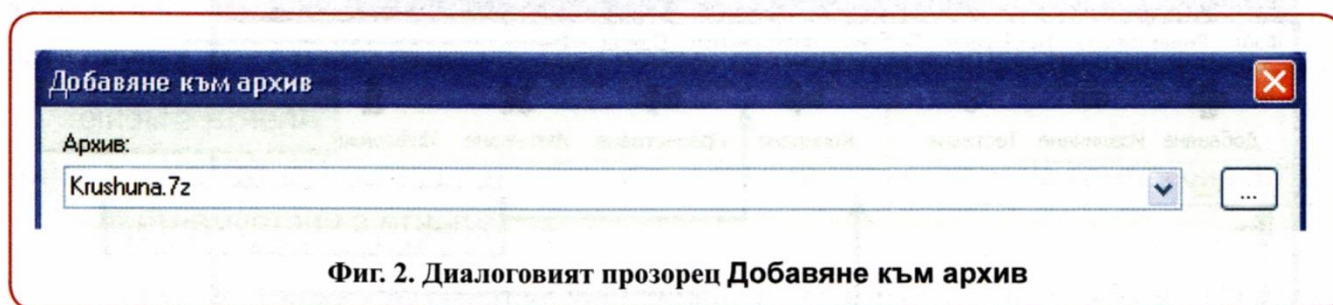
не (Edit), Показване (View), Любими (Favorites), Инструменти (Tools), Помощ (Help). Под нея е лентата с инструменти, с помощта на които се осъществяват следните операции:

- Добавяне – добавя файл към съществуващ архив или създава нов архив;
- Извличане – разархивира част от архив или цял архив;
- Тестване – проверява създаден вече архив;
- Копиране – копира файлове от и в съществуващи архиви;
- Преместване – премества файлове от и в съществуващи архиви;
- Изтриване – изтрива един или няколко файла от архив;
- Информация – дава подробна информация за даден архив, файл или папка.

За всяка операция, файлът или папката трябва да са маркирани. За пример да създадем архив. В кореновата папка на устройството C:\ създайте папка Архивирание и копирайте в нея папката Razdel 3Krushuna от диска към учебника. Отворете списъка на комбинирания кутия, означена на *Фиг. 1* с път на текущата папка. Щракнете двукратно с левия бутон на мишката върху реда Компютър от списъка. Изберете последователно

устройството C:\, папката Arhivirane и папката Krushuna, за да видите файловете, които ще архивираме. Върнете се обратно в папката Arhivirane с бутона за връщане в предходна папка, показан на *Фиг. 1* (или с натискане на клавиша BackSpace).

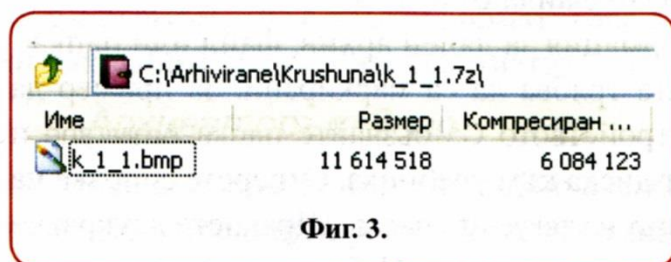
Маркирайте папката Krushuna и натиснете бутона Добавяне. Отваря се диалоговият прозорец Добавяне на архив. От списъка на комбинираната кутия Архив изберете предлаганото от програмата име Krushuna.7z (по премълчаване, програмата предлага за име на архива, името на папката, която ще архивирате) или напишете в полето избрано от вас име на архив, като разширението му трябва да е 7z (*Фиг. 2*). От бутона , намиращ се вдясно на полето Архив може да се избере папка, различна от текущата, в която да се запише архивът. Натиснете бутона ОК и ще се отвори прозорец, който визуализира процеса на архивиране.



След приключване на архивирането, в работното поле на прозореца ще видите, освен папката Krushuna, и файла-архив Krushuna.7z. Маркирайте папката и натиснете бутона Информация. Запомнете размера на папката, показан в страницата General на диалоговия прозорец (трябва да е 108 MB). Направете същото и с файла-архив. Каква е получената степен на компресия? Сравнете получения от вас резултат с оценката на програмата (Ratio), дадена в страницата Archive на диалоговия прозорец.

За да разгледате съдържанието на архива, щракнете два пъти с левия бутон на мишката върху него. Появява се папката Krushuna, но обърнете внимание на пътя – C:\Arhivirane\Krushuna.7z\. Програмата е отворила архива и показва, че той се състои от една папка. Отворете тази папка. Показват се файловете, които са архивирани заедно с нея. Програмата 7-zip, както и Windows Explorer, дава възможност да сортираме файловете в архива по Име, Размер и т.н., като натиснем съответния бутон в лентата.

Когато с програмата 7-zip архивираме няколко файла едновременно, не може да се види степента на компресия на отделните файлове. За да можем да видим степента на компресия на един файл, трябва да е добавен в архива самостоятелно. Върнете се с натискане на клавиша BackSpace назад, докато се покаже файлът



Krushuna.7z, маркирайте го и го изтрийте, с натискане на бутона Изтриване или клавиша Del. Отворете папката Krushuna и маркирайте файла k_1_1.bmp. Архивирайте го. Направете същото и с файла k_1_2.jpg. Когато отворите файла k_1_1.7z (*Фиг. 3*), ще видите, че

размерът на файла k_1_1.bmp след компресирането е паднал от 11 МВ до 5943 КВ. Пресметнете каква е приблизително степента на компресията.

Сега разгледайте файла k_1_2.7z. Сравнете размера на архивирания файл с размера на оригиналния k_1_2.jpg. Стана нещо странно – вместо да се намали, размерът на архивирания файл се увеличи! Причината е, че графичните изображения, съхранени във формат JPG, са компресирани, а във формат BMP – не са. Затова **няма смисъл да се архивират** отделни файлове с разширение JPG. Подобен ефект – невъзможност за компресиране, може да се наблюдава и при файловете, съдържащи изпълними програми – с разширения EXE и COM. Но причината там е съвсем друга – отсъствието в такива файлове на условия за компресиране (дълги последователности от еднакви байтове или често срещани се редици от едни и същи байтове). Естествено, никаква компресия няма да получим и ако се опитаме да архивираме файл, който е архив, създаден с друг архиватор.

Речник

archive	ъркайв	архив
archiving	ъркайвинг	архивиране
compressing	къмпрèсинг	компресиране
losses	лòузис	загуби
lossless	лòслес	без загуби
ratio	рèйшоу	пропорция, коефициент

Въпроси и задачи

- Има ли смисъл да се използва архивиране, ако:
 - се налага да изпратите по електронната поща файл с размер 1 КВ;
 - твърдят ви диск е почти пълен и трябва да освободите място, за да може оперативно да записвате данни на него;
 - от последната екскурзия на класа сте заснели филм с продължителност 1 час. Искате да го копирате на DVD и да го раздадете на вашите съученици;
 - ваш познат иска да му запишете, колкото може повече, песни от вашата колекция, за да ги прехвърли на своя компютър;
 - за презентацията си, сте намерили илюстрация във формат TIF с размер 100 МВ и трябва да я прехвърлите с флаш-памет на училищния компютър;
 - имате 100 илюстрации във формат JPG, всяка с размер около 1 МВ и трябва да ги прехвърлите с флаш-памет на училищния компютър.
- Посочете файлови формати, за които компресирането няма ефект. Обяснете причините.
- Компресируйте с алгоритъма от урока следната последователност:

7 7 3 8 8 8 8 8 2 3 4 4 4 5 6 6 6 6 6 6 6 6 4 4 4 4 4.