



Поиск и удаление дублирующейся информации

Эпоха дубликатов

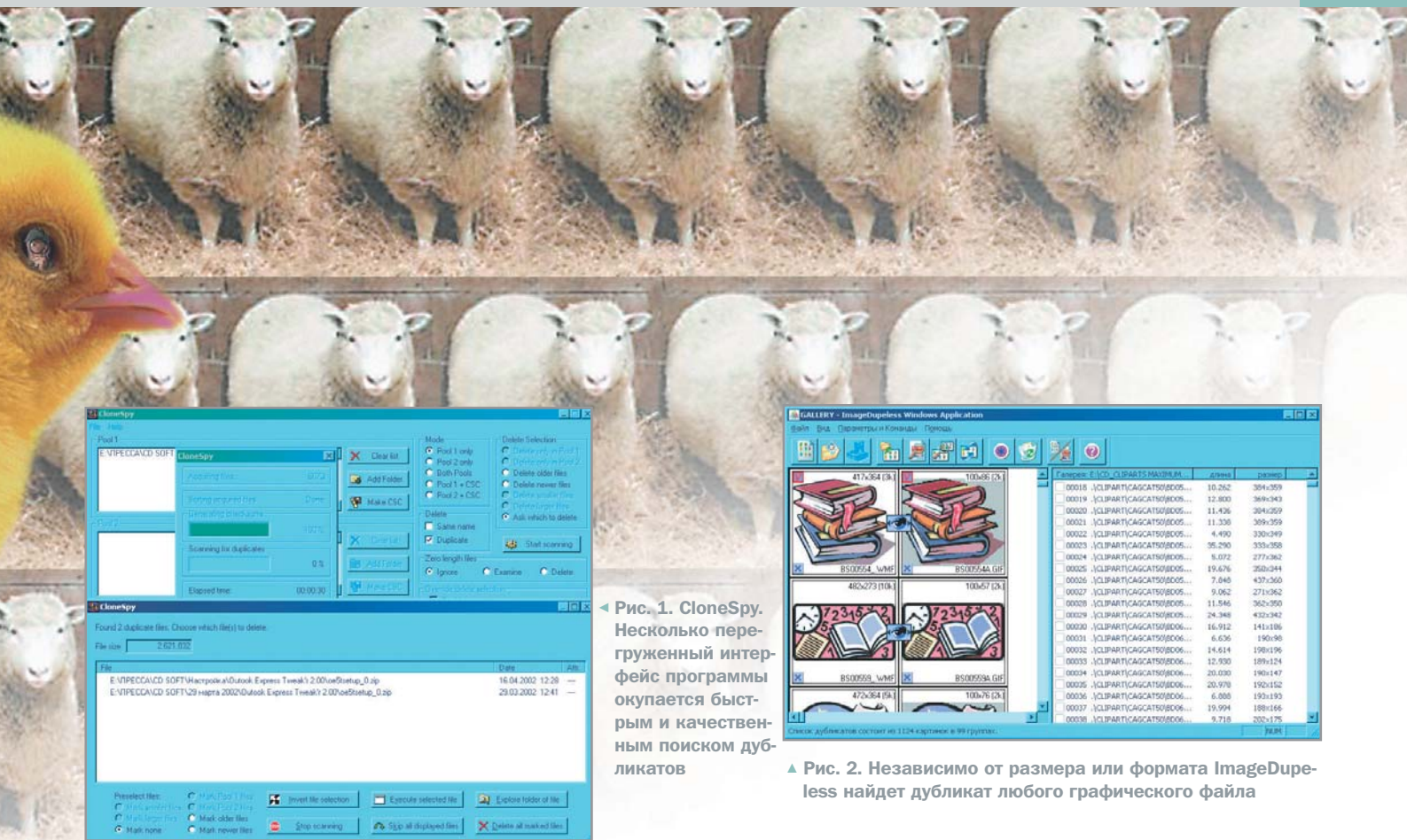
Сегодняшнее положение дел в сфере всеобщей информатизации многих ставит в тупик: информации слишком много для правильного ее восприятия и обработки. Растущие словно после дождя многочисленные тематические сайты, сайты с архивами freeware/shareware программного обеспечения приводят к ожидаемым последствиям: в компьютере рядового пользователя с каждым днем растет объем дублирующейся информации, что приводит к «замусориванию» винчестера.



«Замусоривание» компьютера дублирующейся информацией начинается буквально с колыбели — с установки операционной системы. Процесс происходит в два раза быстрее, если за компьютером работают несколько пользователей: одни и те же шаблоны документов, создание своих коллекций музыки, копий документов и программного обеспечения. Пользователи начинают интересоваться проблемой, когда свободного места на винчестере остается до-

статочно лишь для относительно стабильной работы системы. В том, что эта проблема решаема, уверены разработчики специализированного программного обеспечения, направленного на поиск похожих файлов различных форматов — исполняемых, графических и музыкальных файлов, текстовых документов. Предлагаем вашему вниманию наиболее интересное программное обеспечение для поиска и удаления дубликатов данных с жесткого диска.





▲ Рис. 1. CloneSpy. Несколько перегруженный интерфейс программы окутается быстрым и качественным поиском дубликатов

▲ Рис. 2. Независимо от размера или формата ImageDupeless найдет дубликат любого графического файла

CloneSpy 1.51

ПЛЮСЫ/МИНУСЫ

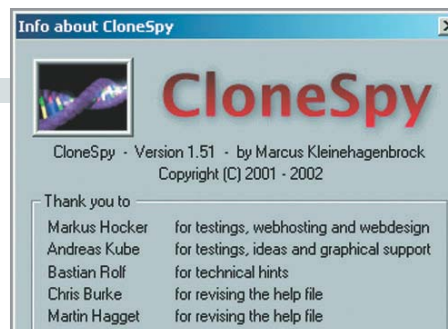
- + быстрый и качественный поиск
- + сравнение двух директорий
- + журналирование всех операций
- + не требует инсталляции

Насколько часто вы скачиваете файлы из Интернета? Ваш винчестер просто переполнен различной информацией, и вы не в состоянии в ней разобраться? Вы хотите «почистить» диск и удалить дубликаты? Попробуйте CloneSpy — уникальную утилиту для поиска дубликатов скачанных программ, находящихся на жестком диске. Программа выполнит за вас всю рутинную работу: освободит место на винчестере, обнаруживая и удаляя дубликаты программ, причем последние будут определены независимо от их имени, даты, времени и месторасположения. Аналогично находятся файлы, имеющие разный размер и содержание, но одинаковое название, что очень часто встречается, когда у пользователя хранятся разные версии одного и того же программного продукта.

Главное окно программы (рис. 1) содержит все необходимые модули для поиска дубликатов.

Так, называемые модули поиска (Pool 1, Pool 2, Checksum (CSC) files) предназначены для поиска дубликатов в двух разных директориях, сравнения двух директорий (пример: коллекция софта на CD-ROM и коллекция, переписанная на винчестер) с целью выявления новых файлов. К тому же CloneSpy способна формировать файлы контрольной суммы из файлов на CD, с тем чтобы можно было определить, нет ли таких же файлов на винчестере. Из дополнительных возможностей программы необходимо отметить следующие:

- ▶ быстрая скорость сканирования выбранных директорий;
- ▶ ведение журнала — файла с записью о найденных дубликатах;
- ▶ возможность выбора между полным удалением и переносом дубликатов во временную папку программы;
- ▶ поиск дубликатов по маске;
- ▶ удаление одного из дубликатов прямо из окна CloneSpy.



Программа производит действительно качественный поиск дубликатов, работать с ней просто и удобно, разработчик дает лишь один совет: если не желаете «снести» систему, не сканируйте папки Windows и Program Files, даже если там (о, ужас!) записаны архивы с программами. Лучше их скопировать на другой локальный диск и продолжить работу. Утилита не требует обычной процедуры инсталляции.

CloneSpy 1.51

Разработчик ▶ Marcus Kleinhagenbrock

Сайт компании ▶ www.clonespy.net

Условия распространения ▶ freeware

Интерфейс ▶ английский

ОС ▶ Windows 9x/Me/NT/2000/XP

Размер дистрибутива ▶ 327 Кбайт

ImageDupeless 1.3

ПЛЮСЫ/МИНУСЫ

- + поиск по содержанию файла
- + возможность работы с очень большими коллекциями файлов
- + работа с иконками файлов без самих файлов
- необходимость предварительного создания образа коллекции

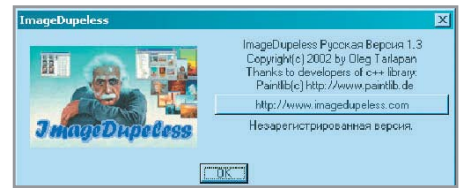
В отличие от предыдущей утилиты, ImageDupeless предназначена для поиска дубликатов изображений в больших коллекциях. Программа поддерживает работу с графическими файлами форматов JPEG, GIF, BMP, PCX, PNG, TIFF, PGM, WMF, EMF, EPS, PSD. Аналогично CloneSpy, сравнение производится на содержание изображений независимо от их размера, формата, разрешения и т. д. Процесс поиска дубликатов заключается в создании коллекции изображений, в ходе чего производится сканирование заранее указанных носителей данных и выявляются изображения, имеющие схожее наполнение (рис. 2). Пользователь должен указать максимальный уровень различия, при котором картинки еще считаются одинаковыми (в процентах; 0% — это очень похожие картинки). Результатом сравнения является набор пар или групп изображений, мера схожести которых меньше предела, заданного пользователем.

Поскольку в галереях с большим количеством изображений, исчисляемым десятками тысяч, использование такой функции вряд ли рационально, то второе предназначение программы — каталогизировать имеющиеся коллекции изображений для дальнейшего сравнения вновь поступаю-

щих картинок с уже сформированной коллекцией (галереей) без полного пересканирования изображений в самой галерее. Это позволяет значительно сэкономить затрачиваемое на всю процедуру время. При этом в процессе сравнения файлы картинок самой галереи могут отсутствовать (храниться на внешнем носителе — CD, например), в работе используется сформированный ранее файл галереи.

Вот дополнительные возможности ImageDupeless:

- ▶ внутренние средства сравнения, удаления (в том числе групп картинок), перемещения и обновления файлов, изменения привязки корневого каталога галереи;
- ▶ новая опция «Сохранять иконки в галерее», предназначенная для желающих использовать галерею в офлайне. С одной стороны, ее наличие позволяет в дальнейшем просматривать состав галереи в виде иконок, не имея исходных файлов с изображениями на диске. С другой стороны, ее выключение в два раза ускоряет процесс создания галереи и в три раза уменьшает размер галереи на диске!
- ▶ автоматическое удаление дубликатов в Корзину;
- ▶ возможность использовать встроенное средство просмотра иконок по двойному щелчку мыши;
- ▶ добавлена возможность запуска встроенного средства просмотра для текущего изображения по клавише «пробел»;
- ▶ дополнительные наглядные информационные элементы, позволяющие быстро



судить о соотношении качества сравниваемых картинок.

Учитывая то, что программа распространяется условно-бесплатно, следует отметить следующие ограничения, присутствующие у незарегистрированной версии этой программы:

- ▶ размер галереи ограничен 2000 изображениями;
- ▶ работа в панели дубликатов возможна только с первыми пятью парами изображений;
- ▶ при работе с окном появляется сообщение об использовании демоверсии программы, функциональность сравнения двух изображений ограничена.

Текущая версия программы доступна для скачивания на сайте разработчика в двух версиях: с английским и русским интерфейсами.

ImageDupeless 1.3

Разработчик ▶ Олег Тарлапан

Сайт программы ▶ www.imagedupeless.com

Условия распространения ▶ shareware, 150 руб.

Интерфейс ▶ английский, русский

ОС ▶ Windows 9x/Me/NT/2000/XP

Размер дистрибутива ▶ 377 Кбайт

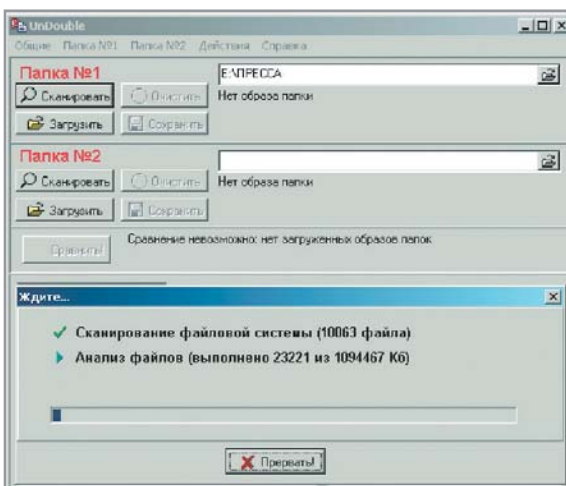


Рис. 3. UnDouble — быстрый качественный поиск обеспечивается за счет создания образов папок

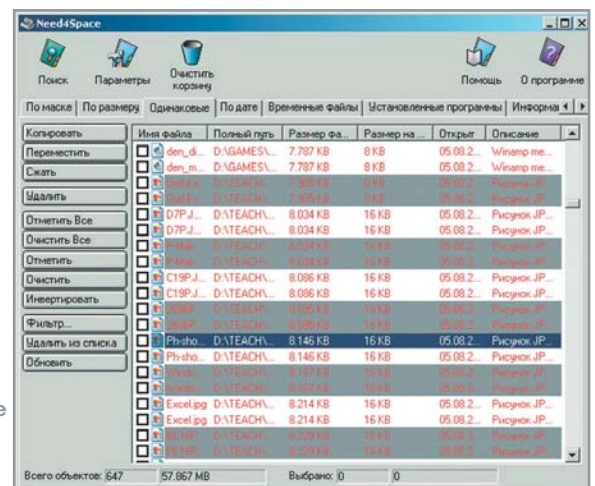


Рис. 4. Need4Space позволит удалить не только дубликаты, но и временные файлы

UnDouble 1.6

ПЛЮСЫ/МИНУСЫ

- + поиск по содержанию независимо от названия и расширения
- + сравнение и поиск дубликатов в двух разных папках
- + возможность восстановления испорченных файлов
- необходимость предварительного создания образа сканируемой папки

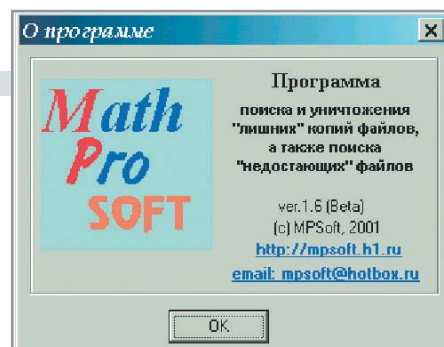
Программа умеет искать одинаковые файлы на диске или в какой-то папке, даже если эти файлы были переименованы, перенесены в другое место, если изменилась дата их создания и т. п. С требуемыми папками утилита работает следующим образом: для того чтобы найти дубликаты, необходимо предварительно создать так называемый образ папки, то есть упорядоченный список всей информации, хранящейся в папке (рис. 3). Причем образы папок можно сохранить на диске и работать с ними в дальнейшем по мере необходимости. Кроме того, для каждого файла программа хранит еще и контрольную сумму,

зависящую от размера файла. Как объясняют разработчики, вероятность ошибки при поиске дубликатов составляет менее одного к миллиарду.

Основные возможности программы таковы:

- ▶ поиск повторяющихся (дубликатов) файлов;
- ▶ сравнение и поиск дубликатов в двух разных папках;
- ▶ поиск файлов, измененных какой-либо программой (шаблоны документов, базы данных);
- ▶ возможность восстановления испорченных файлов любого программного продукта путем создания образа папки и сравнения его с образом папки с работающим программным обеспечением.

Для поиска дубликатов можно использовать любой из следующих типов: поиск по контрольным суммам файлов, поиск с учетом имени файлов, а также поиск с полным сравнением содержимого (в зависимости от объема анализируемых данных



операция может длиться довольно долго).

Результаты сравнения отображаются в главном окне программы, просмотрев которые, можно удалить найденные дубликаты в Корзину. Можно также открыть папку с помощью Проводника для просмотра или запустить файлы.

UnDouble 1.6

Разработчик ▶ MPSoft
Сайт ▶ <http://mpsoft.h1.ru>
Статус ▶ freeware
Интерфейс ▶ русский
ОС ▶ Windows 9x/Me/NT/2000
Размер дистрибутива ▶ 480 Кбайт

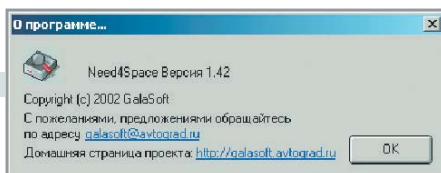
Need4Space

ПЛЮСЫ/МИНУСЫ

- + многоцелевая направленность на удаление разного «мусора»
- + поиск дубликатов по заданным маскам
- маловато дополнительных возможностей

Need4Space — многофункциональная утилита, среди возможностей которой не только работа по поиску и удалению дубликатов файлов, но и удаление временных файлов и файлов, оставшихся после удаления различного ПО, очистка Корзины, доступ из программы к апплету «Установка и удаление программ» и т. д. Программа производит поиск дубликатов различными способами:

- ▶ поиск по заданным маскам (*.TMP, *.BAK);
- ▶ поиск файлов по времени последнего открытия;
- ▶ поиск файлов по заданному размеру файла (больше или меньше заданного значения);
- ▶ поиск одинаковых файлов вне зависимости от имен и расширений файлов.



Интерфейс программы достаточно прост, интуитивно понятен (рис. 4) и не вызывает трудностей в работе. Программа позволяет произвести поиск при помощи заранее настроенных фильтров, после чего вы можете полностью удалить найденные дубликаты или перенести их в заданную папку или в Корзину.

Немного несправедливо, но функция сжатия найденных дубликатов доступна только для пользователей, работающих в ОС Windows NT4/2000.

Need4Space 1.42

Разработчик ▶ GalaSoft
Сайт ▶ <http://galasoft.avtograd.ru/need4space.html>
Условия распространения ▶ freeware
Интерфейс ▶ русский
ОС ▶ Windows 9x/Me/NT/2000/XP
Размер дистрибутива ▶ 377 Кбайт

Закключение

Меня порадовала практическая реализация программ: с помощью вышеописанного софта на моем винчестере было найдено много продублированного софта и графики, в основном скачанных из Интернета и портящих мне настроение своим постоянным клонированием.

На мой взгляд, наиболее удобными являются утилиты CloneSpy и ImageDupeless. Хотелось бы отметить, что описанный здесь софт, излечивая «замусоривание» винчестера, никоим образом не влияет на причину этой «болячки». А проистекает она, очевидно, от бессистемного складирования данных. Вот пара практических советов по профилактике разведения дубликатов.

- ▶ Организуйте в личном каталоге логичное дерево папок, так чтобы их содержимое явно следовало из их имени.
- ▶ Для уменьшения количества копий текстовых файлов пользуйтесь функцией хранения версий документов в текстовых процессорах MS Word, WordPerfect и OpenOffice.org Writer.

■ ■ ■ Сергей Уваров