



Care and feeding of disk space on Linux

Cleaning House

You wouldn't believe how much of the disk space on your system is wasted, filled with duplicate files, or cruft you don't need. *By Joe "Zonker" Brockmeier*

JOE "ZONKER" BROCKMEIER

Joe "Zonker" Brockmeier is a freelance technology journalist, editor, community builder, and Software Freedom advocate. Brockmeier was the openSUSE Community Manager from 2008 to early 2010. Brockmeier now writes for *Linux Pro Magazine* and a number of other technology publications. His personal site can be found at <http://www.dissociatedpress.net/>, and you can follow him on Twitter and Identi.ca as @jzb.

The sad fact is, no matter how much disk space you have, it starts filling up pretty quickly with unnecessary junk. And necessary junk. A few thousand MP3s, lots of work documents, some movies you've downloaded (legally, of course) or ripped for watching on that lovely new tablet you picked up Pretty soon that vast expanse of empty bit-bucket has been given over to monstrous amounts of data. How does that happen?

The fact is, it's not all your fault. A lot of stuff happens behind the scenes. Programs cache data and suck up disk space without your noticing. Logfiles grow and grow. And you contribute too, of course, saving everything because you might just need it "someday."

I'll take a look at a few tools on Linux you can use to reclaim control of your disk space and help deal with the digital overflow.

Getting Started

Because Ubuntu, Debian, and Linux Mint users make up a pretty huge share of the Linux "market," I'm going to concentrate on tools you can find in the most recent Ubuntu and Mint releases. Specifically, I'm working with a Linux Mint 11 desktop. Your favorite distribution might not have quite the same packages I will refer to, but they're all open source, and you should be able to compile them if they're not already in your distro's repositories.

What kinds of tasks do you need to do to clean up disk space? You could go through your directories by hand and remove files one by one. In the old days, I spent more time than I care to think about searching hard disks using `find` and other utilities, trying to find the disk-slurping culprits.

You can do amazing things with `find`, but there're much easier ways to find obsolete, duplicate, and unnecessary files – starting with BleachBit, a utility that not only frees disk space but also helps guard your privacy.

Clean Junk; Maintain Privacy

BleachBit [1] is a Linux and Windows app licensed under the GPLv3. This fairly simple application can make your life much easier very quickly.

BleachBit scours your hard disk for files you don't need – and might not want. For example, it looks for the cache files from Adobe Reader, APT cache, logs from Audacious, Beagle's search indexes and logs, and so on. It also scours your system for browser cookies, cache, and the like. It will also do a "deep" scan for backup files, temp files, and extraneous `.DS_Store` files. (Even if you're not a Mac user, if you've downloaded stuff that's been created on a Mac, you can wind up with these littering your disk.)

To start, you'll need to grab the bleachbit package, which is a piece of cake on Mint and other Debian-based systems:

```
sudo apt-get update
sudo apt-get install bleachbit
```

The update command is really only necessary if you haven't done that recently.

BleachBit has two modes: regular user and superuser. For some reason, at least on Linux Mint, the *BleachBit (as root)* menu option just does nothing, so you'll want to launch it from the terminal with `sudo`, assuming you *do* want to run it with superuser privileges so you can do things like zap the extra files that APT brings in that you don't need.

Once you've started the application, you'll see something similar to Figure 1. See? It's not scary at all. Basically you have a couple of options. First, you see on the left-hand side a set of applications and the files you can choose to work with. Check the radio boxes next to the applications and file types that you want to target. If you click on a specific app or type of file, BleachBit will give you a short description of that file.

The two options are *Preview* and *Delete*. I'm pretty sure you don't need a lot of guidance here. Be sure to preview what BleachBit is going to do before going ahead and hitting *Delete* to avoid any nasty surprises. When you run the preview, it will look a lot like Figure 2, except of course you should have different files on your system.

What kind of crufty nastiness does BleachBit target? It looks for things like backup files, such as those that end in `~` or `.bak`. BleachBit does not have super-magical powers to peer into files and divine whether you actually *need* any of those files or whether they're useful. It simply operates on the basis of common file names that applications use or directories that are typically used for cache, cookies, and other (presumably) disposable files.

Note that if you delete, say, Firefox's cookies, you will find yourself logged out of a lot of services. BleachBit will also perform a "Vacuum" on the Firefox database to save space. Note that Firefox also offers features to delete most of these files, so my policy with BleachBit is to avoid whacking Firefox and other browser data and to concentrate on other files.

One thing that's not terribly obvious from the main screen is languages. BleachBit can delete language support that's not needed from the system as well. (See Figure 3 for the language preferences.) To get to the preferences, go to the Edit menu and then select *Preferences*.

Once you've decided, BleachBit will provide a confirmation after you hit *Delete*. If you're really sure, then go ahead and hit the button and clear out some space. How much space can you reclaim? I got back about 500MB on one system and more on a desktop I've used more regularly. It does depend, of course, on how much cruft you have. You probably won't clear enough space for the entire run of *Star Trek: Next Generation* ripped to disk, but you will likely get enough space back for a couple of CDs worth of MP3s at least.

Finally, you might want to declare some areas of your disk off limits. Go to the Preferences dialog (Figure 4) and add any files or directories you might want to be doubly sure are not deleted.

As you can see, BleachBit can trim your disk space pretty quickly, but what about those things that BleachBit doesn't tackle, like duplicate files?

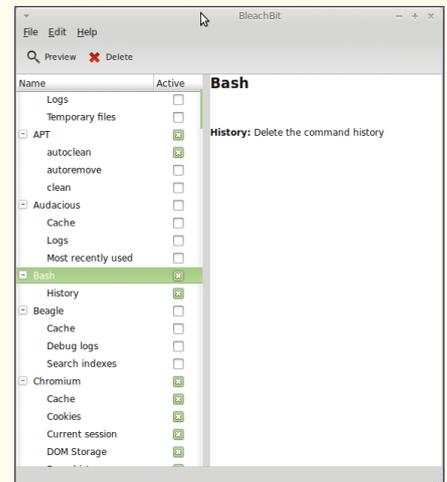


Figure 1: Starting up BleachBit.

FSlint

Short for "Filesystem Lint," this little lovely tool looks for "lint" on your filesystem in the form of wasted disk space, empty directories, temporary files, and (most importantly) duplicate files. To grab this one, get the `fs-lint` package with `sudo apt-get` or check for it in the software manager.

Once installed (Figure 5), you have the option of

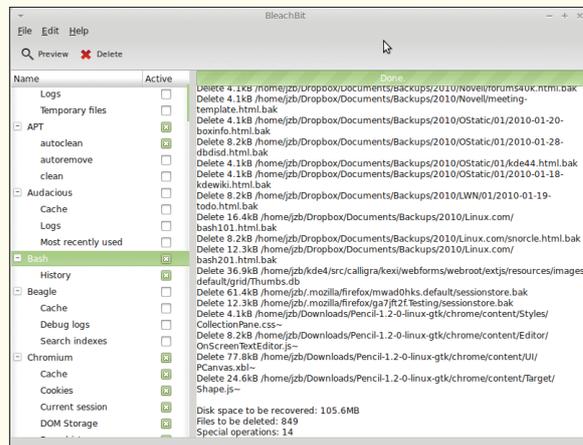


Figure 2: Previewing files to be deleted.

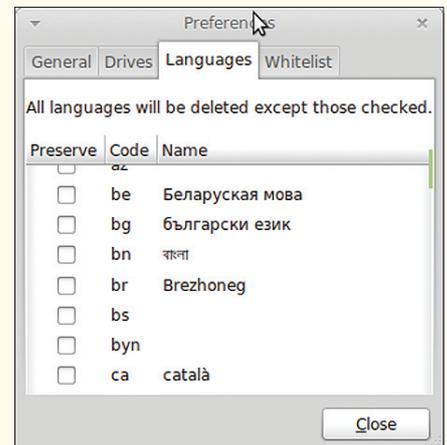


Figure 3: BleachBit Preferences dialog.

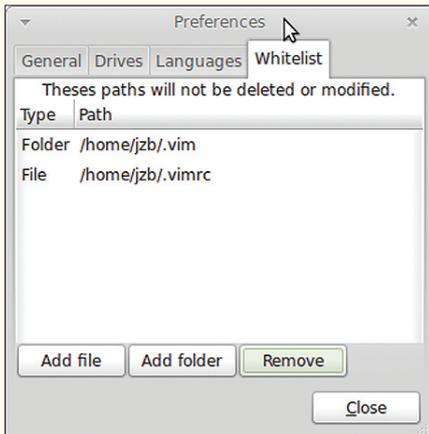


Figure 4: BleachBit Whitelist tab.

choosing the search path and the type of files. The tab on the left-hand side determines the type of search you're doing. Because you're most concerned with duplicates, go ahead and search for them by clicking the *Find* button in the lower left-hand corner. It might take a few minutes, depending on how many files you have, the speed of your system and disk, and so on.

After grinding for a bit, FSlint returns a set of files it believes to be duplicates. Although it's not immediately obvious, you can double-click on files in the FSlint window and open them to see what they are and confirm that they actually are duplicates.

From the FSlint window, you can choose to delete files, save the list to a file, or merge files. *Be very careful* with the merge feature. By default, it will want to merge all files *except* those selected, which is very counterintuitive.

FSlint uses `find` behind the scenes. You can add additional parameters for `find`, paths to exclude, or both, as shown in Figure 6.

If you're looking to save disk space by uninstalling disk hogs, check the *Installed Packages* tab, as shown in Figure 7. As you can see, some of the biggest packages aren't exactly optional. The Linux kernel image package takes up 114MB, Sun Java consumes about 90MB, and Google Chrome checks in at 91MB. You could do without Java or Chrome, I suppose, but good luck doing without the kernel!

Command-Line Fun

The GUI tools are great, but what if you are trying to clean up a system that's not running a desktop? How about scripting?

You could use `find` for a lot of this, but it's easier to use `fdupes`. This utility finds duplicate files when given a set of directories, such as

```
fdupes -r ~/Dropbox
```

which tells `fdupes` to search the Dropbox folder under the home directory for any duplicate files. On my disk, it turns up quite a few, because I've been sloppy managing my e-books.

If you just want to know how much space is being consumed by duplicate files, the `fdupes` "summarize" option (`-m`) just reports the number of duplicate files:

```
fdupes -m -r ~/Dropbox
```

That command, for instance, tells me my system has 221 duplicate files that occupy 87.4MB. Note that it doesn't do anything to take care of those, it just provides information about duplicate files.

The utility also has an option to report the first duplicate in a list of files. Assuming you're confident in the report, you can use this list to remove the duplicate files. Run

```
fdupes -r ~/directory -f
```

and you'll only see the duplicate set of files, not the "original." Of course, you're depending on `fdupes` to determine the "original" from the "duplicates," and it's not possible to tell the utility how to distinguish the file names of duplicates from that of the original.

If you *really* trust `fdupes`, it will delete files for you if you invoke the `-d` (delete) option: It will prompt you for which files to delete, and it will give you a list of files it believes to be duplicates, with a number next to each one. At this point you can choose the number of the file to preserve, or you can choose to preserve all.

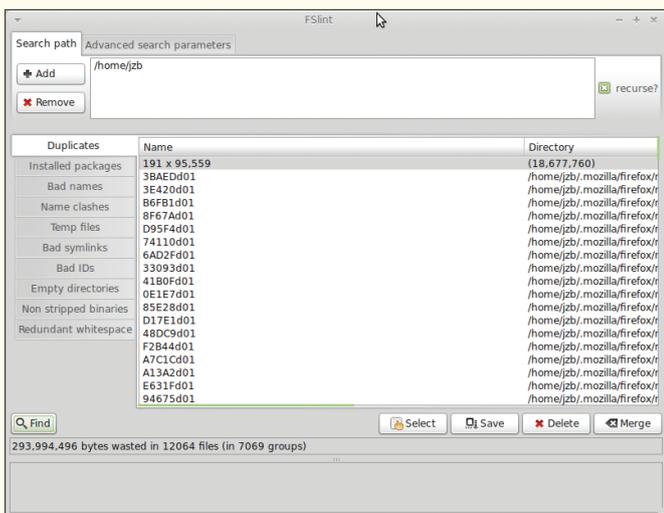


Figure 5: FSlint shows duplicate files.

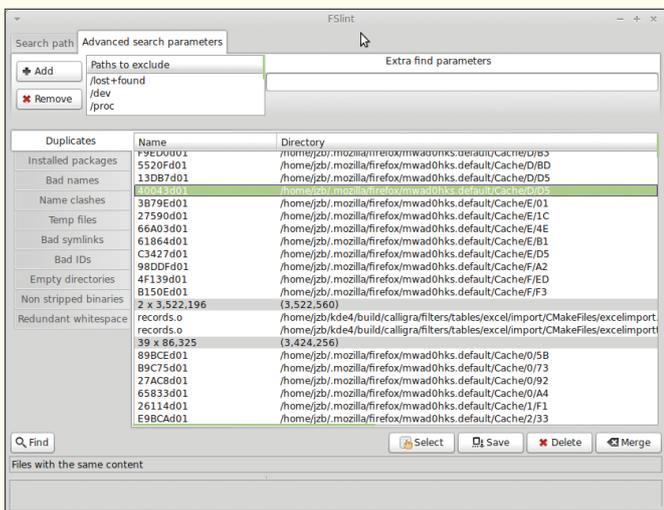


Figure 6: FSlint advanced search.

You can even go one step further with the `-N` (no-prompt) option and let `fdupes` do its thing unaided. I'd be careful with that one, though.

As with the other tools, just install the `fdupes` package with your favorite utility for managing packages.

A Quick Word on JDiskReport

Another tool worth checking out is JDiskReport [2]. I'm not covering that tool here because it was recently featured in *Linux Magazine* [3]. JDiskReport is not available via the repos, unfortunately, but you can download the latest version on the JGoodies website [4]. It runs on Linux as a standalone application or in the browser and is also available for Windows and Mac OS X.

JDiskReport shows you how much space is being used on your disk, the size distribution of your files, type distribution of your files, and much more.

Slimmed-Down Disks are Happy Disks

Even with 1TB drives, they're still overloaded. The tools highlighted in this article are very useful in trimming the fat. The last resort, of course, is off-loading some of that fat to external storage or cloud storage to save more space.

Many of the files on your disk are things you want to keep around, but not necessarily things you want all the time. How do you manage those? That's what I'll look at next month. Until then, get rid of those unnecessary files and next month you can work on offloading the files you use infrequently. ■■■

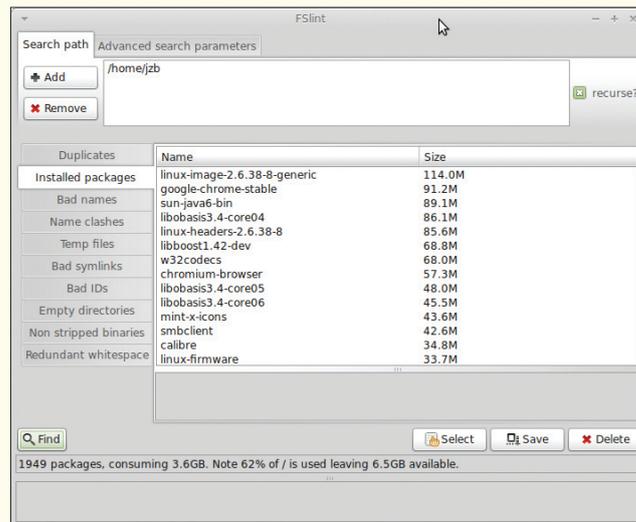


Figure 7: FSInt Installed packages tab.

INFO

- [1] BleachBit: <http://bleachbit.sourceforge.net/>
- [2] JDiskReport: <http://www.jgoodies.com/freeware/jdiskreport/>
- [3] "JDiskReport" by Tim Schürmann, *Linux Magazine*, May 2011, g. 46
- [4] JGoodies: <http://www.jgoodies.com/downloads/index.html>



Online Training at Linux Magazine Academy

Monitoring with Nagios

Getting your IT under control the easy way:

- Web front-end
- Monitoring Windows/Linux/Unix
- Structuring the configuration
- Monitoring SNMP-components
- Nagvis, Grapher V2, and ND02DB add-ons



Many practical examples!