

An up-to-date look at free software and its makers

PROJECTS ON THE MOVE

The Esperantilo editor helps users write in the world language, and Kleansweep searches for dormant files. If you are an expert on compression algorithms, you can win 50.000 Euros. Also, the lowdown on the Cdrecord feud, and confirmation of the Etch release.

BY CARSTEN SCHNOBER AND MARTIN LOSCHWITZ

Fans of Esperanto believe that a common language for the whole world would solve many problems. Esperanto [1] was invented and introduced by Ludwik Lejzer Zamenhof in 1887 as a means of simplifying worldwide communication.

Language Under a Free License

To this day, Esperanto is supported by a small but faithful community, and it is no coincidence that many Esperanto fans are also members of the Linux community. Besides the cosmopolitan goals and worldwide backing, Linux and Esperanto share similar license models. The inventor gave up his rights to the artificial language after its release, thus creating what may well have been the first free license since the invention of copyright.

If you are interested in learning and using Esperanto, Esperantilo [2] is a good choice of editor (Figure 1). It includes spelling and grammar checkers, and it can translate Esperanto into English, German, and Polish. As Esperanto

is designed to be easy to use and to follow regular grammatical rules, machine analysis is easier than with most other languages with their grammatical rules and exceptions.

Texts can be translated either automatically or semi-automatically. In the latter case, users can intervene if the program gets stuck, for example, if a word is not in the directory, or if there is more than one possible translation. In automatic mode, the editor highlights unknown words in the translated text for the user to edit.

The Esperantilo homepage has executables for Linux and Windows; appropriately for the free language and the free operating system, the editor is a free software tool licensed under the GPL. If you intend to use the internal translation tool, you need both the editor and the dictionary, which is available from the same download site.

Meta Search

Search engines are an integral part of any Internet user's daily life. If you prefer not to blindly trust what one search

engine tells you, you can crosscheck the results with those of a competitor's engine. To avoid wasting too much time searching and crosschecking, you might prefer to use a meta search machine such as Metacrawler [3].

The Pinot [4] front-end makes searching across multiple engines even simpler. Pinot is a desktop program that uses the GTK interface and integrates all the major search engines (Figure 2). Users can select their preferred search engines and enter a search key. Pinot then retrieves the results from the selected offerings and displays them in the *Live query* tab.

Besides web search engines, Pinot will also query news services, Freshmeat – the online software catalog, and the source code database at Koders.com. You can add custom web services, provided they can be queried using the SOAP API.

Pinot includes a web browser to display the search results. Selecting *View* in the drop-down menu for a match will open the page for browsing in the *View* tab (Figure 3). However, the minimalist

browser does not let users type in URLs.

Pinot also supports searches against the local filesystem, although users are required to manually create an index. The developers do not view Pinot as a competitor to desktop search engines such as Beagle: instead Pinot provides a front-end for any kind of search engine.

Clean Home

Despite the best laid plans of package management, there can be no overlooking the garbage that clutters up a typical desktop. The home directory being the major candidate for regular cleanups if you want to avoid obsolete configuration files eating up your hard disk. Many users will have written their own shell scripts at some time to help them clean up. Kleansweep [5] combines several mechanisms for locating unwanted files, adding a pretty neat GUI into the bargain (Figure 4).

Kleansweep searches either the directories specified by the user or the whole file system, and gives users the ability to specify the search criteria. The program will find empty files and directories, orphaned symbolic links, thumbnails, and backup files like the ones that many editors create.

After completing the search, Kleansweep displays a list of the files and directories it has found, sorting the list in tabs that indicate why it thinks the entry is obsolete. Users can then select the entries they want Kleansweep to do away with.

The program then goes on to offer to store the selected files in a compressed archive before deletion; this gives you the ability to undo any damage caused by accidental file removal. You can then remove the backups after a definable period if you are sure that you can really do without the files in the archive.

Text Compression

With data volumes continuing to grow to match ever increasing hard disk sizes,

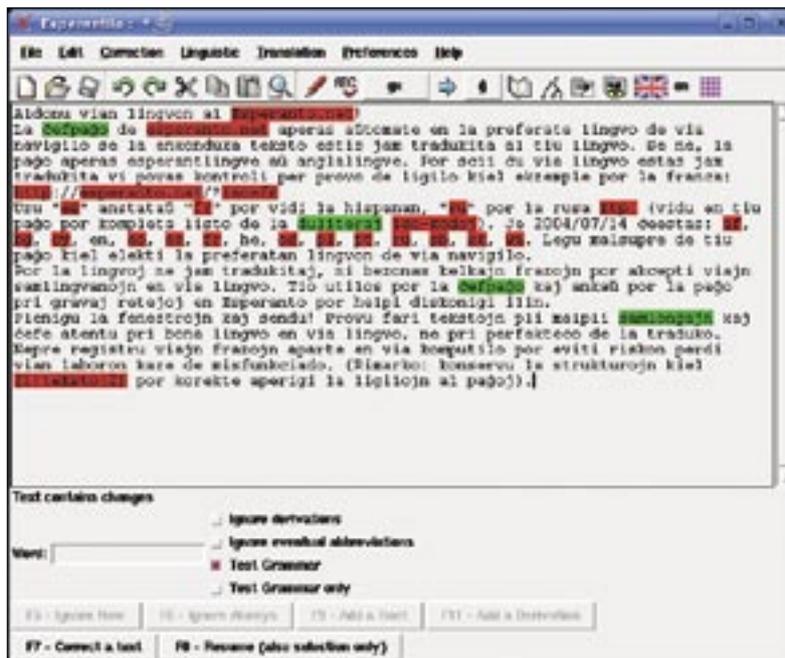


Figure 1: The Esperantilo editor has a spellchecker for the world language, Esperanto.

efficient compression plays as important a role as ever. MP3 takes the crown as the best-known compression algorithm with its ability to compress audio files to support file transfers over normal Internet connections. Transferring raw audio data would consume enormous amounts of bandwidth, and would be unthinkable considering the sheer volume of audio data flying through the web today, although I suspect the music industry might take a different view to this.

Compression algorithms are at their most efficient when designed to handle a specific file type. For example, MP3

simply discards frequencies that are (virtually) inaudible to humans to save space. Applying this technique and others helps MP3 shrink an audio file by about 90 percent. Traditional methods support the space-saving storage of text files based on compression techniques that focus on repetitions. For example, Zip achieves compression rates of about 70 percent.

ASCII files are good candidates for compression, however, there are just a few compression tech-

niques for text. The compression factor could be improved by applying algorithms that analyze the text content. This is an area of research that borders on artificial intelligence. Of course, text compression techniques can't simply leave out information, in contrast to MP3's lossy compression.

To motivate developers to work on advanced text compression algorithms, Marcus Hutter is offering an award of 50.000 Euros to the programmer who compresses a text file better than any predecessor [6]. The test file for this compression contest comprises the first

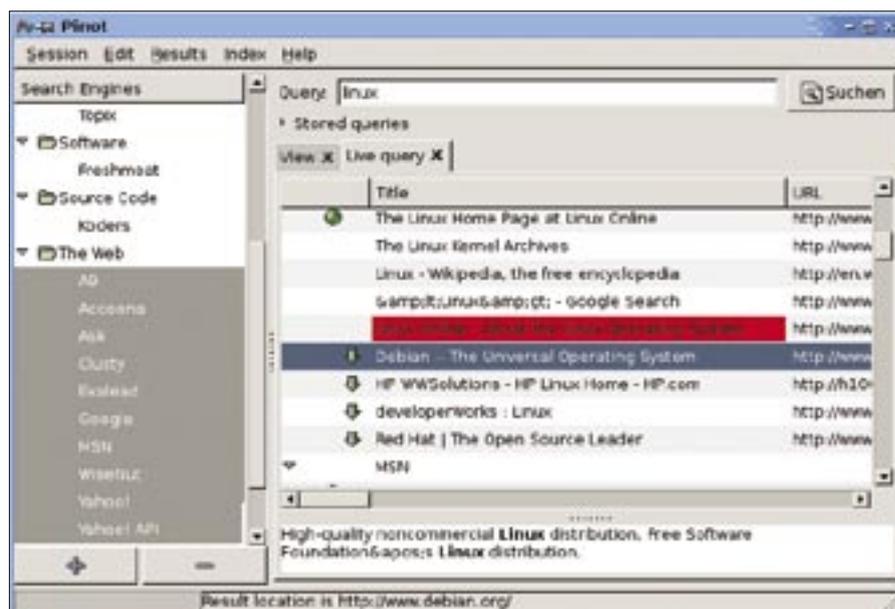


Figure 2: Pinot offers a desktop interface to the major search engines.

100 Mbytes of Wikipedia from March 3, 2006, which explains the competitions motto of “Compressing Human Knowledge.” The current leader had achieved a compression ratio of about 82 percent when this issue went to press.

The Cdrecord Feud

The feud between Debian project representatives and Jörg Schilling has been going on for months. The bone of contention is Schilling’s popular Cdrecord program. Schilling has complained that Debian distributes a modified version of his software without notifying users of the fact, and he states that this is damaging to his reputation.

Among other components, Debian developers have contributed a number of patches for handling SCSI drives on kernel 2.6.x. And the project is in some doubt concerning the compatibility of the Cdrecord license, the CDDL (Common Development and Distribution License), with DFSG (Debian Free Software Guidelines).

Mediation Failed

In August, Debian’s Joerg Jaspart tried to mediate [7]. Jaspart had been the co-maintainer of the Cdrecord package along with Eduard Bloch for a couple of versions. His suggestion was to fork the code base and create a kind of Debian-burn tool, featuring everything the tool needed for Debian.

Just a short while later, Schilling launched a counterattack [8], demanding the dismissal of Eduard Bloch, stating “He has been the biggest problem for Debian in the past years.” Schilling went on to criticize the fact that Debian had modified the source code, demanding that

Debian remove the changes, and use the unmodified source code because it “does not need any changes in order to work correctly”. But, as we all know, it is not uncommon for distributors to modify the source code, in part to integrate a tool more neatly with the look and feel of a distribution.

Finally, Schilling asks the developers to read the CDDL to settle any issues developers and users might have with incompatibility between the CDDL and the DFSG; at the same time criticizing the FSF GPL FAQ as wrong.

Debian may decide to follow Jaspart’s suggestion and fork a previous version of Cdrecord, but on the other hand Cdrecord is not the only burning tool around, and people in the Debian community are

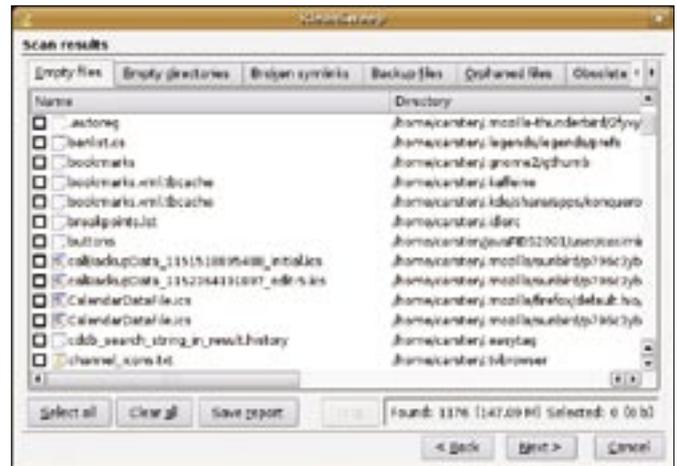


Figure 4: Kleansweep searches the whole filesystem for unwanted files.

starting to ask if a fork is worth all the trouble.

Etch Release Confirmed

Coming as it did so soon after the Sarge release, many people thought that the announcement of a release date for Etch was a joke. But now Debian has confirmed the December 2006 deadline for Debian GNU/Linux 4.0, alias Etch. The bulletin at [9] also publicly confirms that 2.6.17 will be the standard kernel for Etch, that X.org is replacing XFree86 – this has already happened in the testing branch – and that Etch will be available on eleven architectures. AMD 64 is a new arrival, and the Motorola M68k architecture has been dropped – much to the chagrin of its fans. But who knows – maybe it will make its way back into the Etch architecture list some day. ■



Figure 3: Pinot displaying search results in the internal browser.

INFO

- [1] Esperanto: <http://www.esperanto.net>
- [2] Esperantilo: http://www.xdobry.de/esperantoedit/index_en.html
- [3] Metacrawler: <http://www.metacrawler.com>
- [4] Pinot: <http://pinot.berlios.de>
- [5] Kleansweep: <http://linux.bydg.org/~yogin>
- [6] Compressing Human Knowledge: <http://prize.hutter1.net>
- [7] Mail from Jörg Jaspart: <http://lists.debian.org/debian-devel/2006/08/msg00478.html>
- [8] Answer from Jörg Schilling: <http://lists.debian.org/debian-devel/2006/08/msg00484.html>
- [9] Debian press release: <http://www.debian.org/News/2006/20060724>