OUT OF THE BOX

AGAINST IT!

Do you get irritated by "dear colleague" emails laden with Word attachments? Christian Perle shows you how you can take a peek at the document without sacrificing memory to Word

It's not always that easy to shake off the world of Windows. Or is it? With Antiword from Adri J. van Os it is possible – even in the **text console** – to display Word documents so they're easy to read.

Clear

Although Antiword can cope with a great many Word formats, it's still a very compact program at only about 100Kb. For the latest version (0.31), you can get the packed **source text** from *http://www. winfield.demon.nl/index.html*.

To install Antiword you need to unpack the source and compile the program. You then need to copy the files into your **home directory** and into the /usr/local branch of the filesystem.

Word for Word

To test Antiword, I have put my head into the lion's mouth and created a brief Word document (tex_is_best.doc) with Microsoft Word 8.0. This includes headings at various levels, a list and a table. The original display can be seen in Figure 1.

In order to feed Antiword with this Word document, enter in the shell:

```
antiword tex_is_best.doc > tex_is_best.txt
```

The > symbol causes the shell to divert the output of the program into the file tex_is_best.txt, otherwise it would just sail through to the console. Many documents can be converted with a simple **shell script**.

In Listing 1 you can see the text output created by

Antiword. What stands out is that the program identifies the headings as such and creates an appropriate numbering. The justified formatting of the first paragraph is also retained. The table representation could be better though – this is where the text-based Web browser w3m could be put to good use.

Pure text data can obviously be better edited with standard Unix tools such as grep than cumbersome .doc files. In order to filter out all the lines containing the word TeX, you need only pass the Antiword output on to grep. In this case the pipe character l links the two programs:

```
antiword tex_is_best.doc | grep -w TeX
```

Apart from pure text, Antiword can also create data in the page description language PostScript, which can then be displayed or printed out using gv. The invocation looks like this:

```
antiword -p a4 tex_is_best.doc > tex_is_best.ps
```

This output format offers more options for text display – for example, font colours are retained (Figure 2).

With filters

If you've tried out all the different text-based mail programs, you should have finally found your way to the best – which is mutt – and you'll certainly want to call up Antiword directly from this program. In the .mailcap file (which you may have to create from scratch) in your home directory enter:

```
application/msword; antiword %s
```

Straight away mutt displays in the internal viewer all **attachments** with the **MIME** type application/ msword, without you having to worry about

Figure 1: Original display in Word 8.0

Text console In addition to the graphical user interface X there are usually several consoles running on a Linux system in text mode. These are reached from running X using Ctrl+Alt+F1 to F6. You can get back to X with Alt+F7.

Source text The form of a software package which can be read by humans. By translating it ("compiling") with a compiler this is turned into an executable program.

Shell script A text file with shell commands, which are executed automatically one after the other.

Home directory The personal home directory of a user. This is where he or

she ends up after a successful log in or with the command cd with no other parameters.

Attachment The optional file attached to an email. This cannot be transferred in binary form, because non-printing characters would get lost. Base64 has become the most frequently used transfer format for binary attachments.

MIME Multipurpose Internet Mail Extensions. A method for specifying standardised file types. Some examples of MIME types are text/plain (pure text file without formatting) or video/mpeg (MPEG-compressed video stream). MIME is used mostly in mail programs and Web browsers.

Quotes If one replies to an email using the reply function, the cited mail text is distinguished by the mail program by placing quotation marks at the start of the line of the text which you are writing. Most mail programs comply with common sense on the Net, which prescribes the character string "> " (greater-than and space). Nevertheless many graphical mail programs do not provide these characters and mark out the quote by using a different colour or a different font.

Manpage Linux, like all Unix systems, has a sort of online reference manual for the installed programs. This aid is called up using man program name, e.g. man antiword.

---

macroviruses. The file manager Midnight Commander (mc) can be extended in a similar way. Select from its menu Command/ Edit suffix data, or use an editor to open the file ~/.mc/bindings. Now enter the following lines there, and save the change:

```
shell/.doc
Open=antiword -p a4 %f | gv -
View=%view{ascii} antiword %f
```

If you press F3 display, when mc's bar cursor is over a Word file, the internal viewer shows the text output of Antiword. If instead you use the Return key, the file will be converted into the PostScript format and passed straight on to gv.

Specialities

A few of Antiword's useful options should not go unmentioned. With -L, the program creates the PostScript output in sideways format ("landscape"). This can make wide tables easier to read.

The option -w col is relevant for the text output and limits the line length to col symbols. If you want to quote the content of the document in a mail something like -w 75 would be advisable, as this
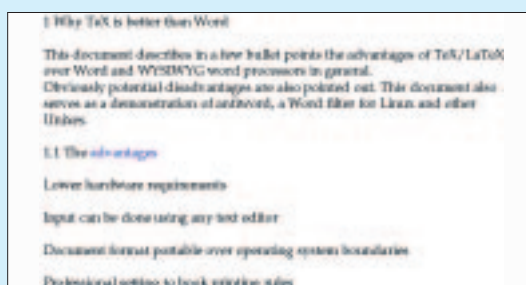
leaves enough space for the **quote** characters.

Text which has been made "invisible" by the Word function of hiding text, is shown by the program when you use the option -s. Whatever else is still hidden in Antiword will be revealed to you by the **manpage**.

Figure 2: PostScript display with gv