

4

SLAY THE WORD AND YOU'LL BE FREE

You hold in your hands something you don't see every day: a book written on a computer without Microsoft Word.

You probably think Word is the stumbling block—the first, last, and final reason why you can't escape Microsoft software. But alternatives exist, and this chapter is all about them. You can use commercial products like WordPerfect, or free software such as OpenOffice.org Writer or AbiWord, or text editors such as TextEdit, which comes with the Mac.

In fact, it was possible to skip Word years ago to produce books that use graphical layouts. Children's books, for example, are typically composed and finished using a page layout program such as Adobe PageMaker or QuarkXPress. Large coffee table books with high-quality photographs are nearly always composed in page layout programs, because only those programs can do justice to reproducing the photographs in a proper graphical layout. Text, formatted or otherwise, has always been the easiest type of data to process. So it stands to reason that it must be possible to *use an alternative to Word*.

And yet, so many people continue to use Word because they need to work with Word documents. Some swear by it, and others swear at it. The application has been around far too long; people should have other choices

by now. “Monopolies become their own worst enemies—particularly in businesses that live or die by technological innovation,” wrote James Gleick in *The New York Times Magazine*. “They get soft. They make poor research choices. They bleed both profit and invention. They poison the marketplace that created them.”*

Microsoft Word Is Not the Final Word

I’m a veteran Word addict, now reformed. Since the late 1980s I relied on Word for all writing and editing. So why do I bite the Word that fed me?

All that time I paid a tithe to the gods in Redmond so that I could eat, paying for upgrades and even migrating to more powerful computers just to run new versions. When the software continued to behave erratically, upgrade after upgrade, I began to feel misused and abused. I had heard rumors that Word harbored viruses, that Word reported back to Microsoft the details of your hardware configuration every time you launched it, and that Word was a manifestation of the devil. I had creepy feelings that somehow Microsoft could read what I was writing.

Of course it’s all Microsoft’s fault. Word’s creepiness is directly related to the random suggestions it throws at you when you least expect them and the way it corrects your grammar before you can finish the sentence. Word reinforces bad rumors and ugly feelings by acting so stubbornly and annoyingly the same as it has since 1995. Yes, if you spend some time, you can eventually figure out how to turn off or adjust some of the most annoying features (see Figure 4-1). But this is *your time* I’m talking about. You already paid for the program; why should you also have to spend more time teaching it to behave?

Let’s not even go into all the bad things Word does, for fear of inducing headaches or even nightmares . . . Like how Word applies the same idiotic default settings to any image you place or table you create, regardless of what came before. Or how, after two decades, I can still make it crash by using the DELETE key to delete characters past the beginning of a paragraph into previous lines of text.

Word is bloated beyond belief. Who actually uses the Data Merge Manager to create form letters? Who customizes Word menus so thoroughly that they can no longer find the spell-checker? How often do you want to choke that paper clip character that pops up with inappropriate suggestions (see Figure 4-2)? How many Word docs have you been unable to open for some reason or another, and when was the last time you got a flawless result by saving in an older Word doc file format?

Nearly 20 years ago, I used Word style definitions to define my book chapters so that they could easily be sucked up into Adobe PageMaker for FrameMaker, only to find that publishers used QuarkXPress, which either ignored or mangled my style definitions, rendering them irrelevant. I cranked up that Microsoft monster just to compose emails, only to find that it converted some of my text into some alien alphabet that others couldn’t decipher.

* Gleick, James. “Making Microsoft Safe for Capitalism.” *The New York Times Magazine*. November 5, 1995.

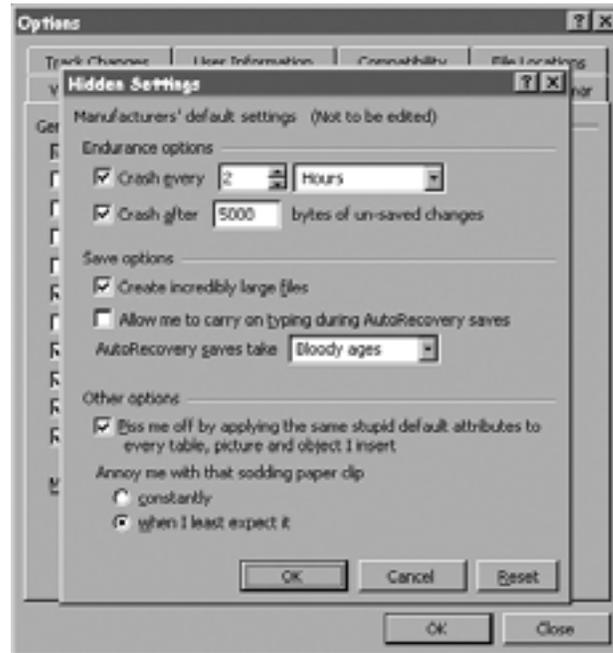


Figure 4-1: Don't you wish you could find these hidden settings in Word? (Thanks to Dr. Norman Clark in the Dept. of Communication at Appalachian State University, Boone, North Carolina.)

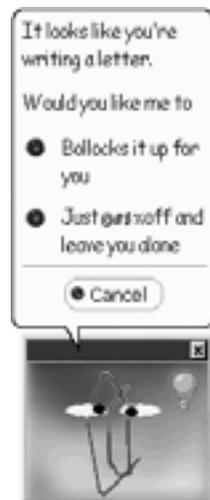


Figure 4-2: Don't you wish this little guy would just go away? Word can be so annoying with these suggestions. (Thanks to Dr. Norman Clark in the Dept. of Communication at Appalachian State University, Boone, North Carolina.)

I tried early alternatives that fell short. I even tried using PageMaker and exporting styled text into Word document files. Ultimately, I had no choice but to continue to use Word. My clients sent me Word documents that I could not open unless I used Word too. I sent Word documents back to them because that's what they wanted. Who knows what evil may have lurked in those Word documents?

What Up, Word Doc? A Format for Disaster

The *Word document format* is the addiction. Not Word itself or some special feature of the program, but the Word doc file format. For a long time, the only way to open a Word doc file was to use Word; without alternatives, businesses migrated to Word and the rest of the Office package and got stuck there. As Word doc files proliferated (and as Microsoft wiped out the competition in word processing programs), they hooked everyone they touched.

Just what is in those files? More than you realize. Word files can violate your privacy. The program is most often configured by default to automatically track and record changes you make to a document. A record of all changes is silently embedded in the doc file every time you save it. It's easy as pie for someone to recover this record and see all the revisions. Most Word document files contain a revision log that is a listing of the last 10 edits of a document, showing the names of the people who worked with the document and the names of the files used for storing versions of the document.

A WEAPON OF MASS DELUSION

Word documents are notorious for containing tracked changes and revisions that could be embarrassing if discovered, and they are easy to discover. The British government of Tony Blair learned this lesson the hard way. In February 2003, 10 Downing Street published an important dossier on Iraq's security and intelligence organizations—the same dossier cited by Secretary of State Colin Powell in his address to the United Nations later that month. It was published as a Word document. A quick examination of hidden revision logs in the Word document revealed that much of the material in the dossier was actually plagiarized from an American Ph.D. student.*

Dr. Glen Rangwala, a lecturer in politics at Cambridge University, discovered that the bulk of the 19-page document was directly copied without acknowledgment from an article in *The Middle East Review of International Affairs* titled "Iraq's Security and Intelligence Network: A Guide and Analysis" (September 2002), written by the student. As a result, during the week of June 23, 2003, the British Parliament held embarrassing hearings on the Blair dossier and other PR efforts by the UK Government leading up to the Iraq war.

* Smith, Richard M. ComputerBytesMan.com. June 30, 2003. See www.computerbytesman.com/privacy/blair.htm. See also Rangwala, Glen. "Intelligence? The British dossier on Iraq's security infrastructure." *The Campaign Against Sanctions on Iraq*. See www.casi.org.uk/discuss/2003/msg00457.html.

And why is the Word doc file format such a moving target? Try to collaborate on a document with a random group of Word users using different versions (Word 97, Word 2000, Word XP, Word 2003, and a couple of Mac versions), and you'll see what a mess you can make with Word doc files. As longtime high-tech pundit John Dvorak pointed out in his *PC Magazine* column,* "It's clear the program is in decline, with too many patches and teams of coders passing in the night. It's about time that it's junked and we get something new. This code can no longer be fixed."

SERVES THEM RIGHT

Ambulance chasers from The SCO Group suing DaimlerChrysler for using Linux forgot to remove tracked changes from a Word doc that identified their previous defendant, Bank of America, and demonstrated the litigation possibilities of Word templates. The press loved it; the SCO Group's credibility was ruined.*

* CNET News.com. March 6, 2004.

The Titanic of Text Editing

You don't want to trust your most important documents to the Word doc file format. One reason is that files are extremely large—it seems like Microsoft stores the design specs and blueprints for the Titanic along with your text. These files are often many orders of magnitude larger than ordinary text files. You can spend a lot of time sending or receiving these files over a network, and they eat up more disk space than they should.

Moreover, Word files are not secure—they can contain code that can destroy your computer. Word lets people create custom programs called *macros* for modifying their Word documents. Excel offers the same feature. But since Word and Excel let you save the macros along with the documents, and since it is possible to hide a virus in the macro code, these documents could easily be turned into Trojan horses carrying viruses. The code can be set to execute as the document opens. In fact, Word documents are now the most common carrier for viruses.

TIP *Don't use Word as your "helper" application for .doc files in Windows. As soon as Word opens a document, your computer could catch a virus—whether you use Word's File menu, click on a Word document on the desktop or in a file list, or click on an email attachment that happens to be a Word doc file. Open your browser's preferences, and set the helper application for .doc files to something other than Word. You might want to register WordPad as the helper application for a Windows system. Otherwise, choose an alternative to Word such as WordPerfect or OpenOffice.org Writer, both described in this chapter.*

* Dvorak, John C. "Kill Microsoft Word." *PC Magazine*. September 7, 2004. See www.pcmag.com/article2/0,1759,1631430,00.asp.

THIEVES USE WORD TO STEAL FILES

Details about one of the largest security holes in Microsoft Word were first published on August 26, 2002 to the popular Bugtraq security list, a service hosted by SecurityFocus, a subsidiary of Symantec.* The security hole has still not been corrected. Essentially, an information thief can steal files from the computer of a person using Word 97.

If you use Word 97 and an unknown person sends you a document to modify, be aware that when you return it, the Word document may contain a hidden copy of files from your computer. The Word document will not be flagged by anti-virus programs. It will also not appear to Word 97 to contain any macros.

The copied files aren't visible in Word, but they are clearly visible using Notepad or Wordpad. The copied files could be documents, Excel spreadsheets, or anything else; they could be located anywhere, even on a secure server. If you have permission to read a file, and you use Word 97 to edit a document from someone who is secretly an information thief, that person could grab the file using "spy" code that can scan for hundreds of files and the INCLUDETEXT field, one of many hidden fields embedded in Word docs. The only way to prevent a file from being stolen is to manually check the fields, which you can find in the document's Properties panel.

If you use Word 97, you shouldn't open and modify a document from someone you don't trust, unless that person will never get the document back. The scheme works best with Word 97, but Word 2000 and 2002 could also be conscripted into service if the attacker can persuade a victim to print the document first.

The security hole is outrageous, yet Microsoft has declined to fix older versions of Word, angering IT professionals. "The only suggestion Microsoft has come up with—examine field codes in your document manually—is so lame I don't know if I should laugh or cry . . . or scream," wrote Woody Leonhard, Certified Office Victim and publisher of the newsletter *Woody's Office Watch*.† "Can you look at a field code and know if it will automatically suck in a sensitive file? How can hundreds of millions of Office users be expected to tell the difference between a safe field code and a spy?"

Microsoft has also angered the community of bug fixers by complaining that the details of this security hole should not have been disclosed without Microsoft's first performing tests. It was only after Woody Leonhard published details in his newsletter that the mainstream press got a hold of the story. Microsoft was forced to make a statement about the problem since the Associated Press was about to release the story to newspapers all over the world. But you have to admire Microsoft's PR machine—the company managed to convince the press that it had disclosed the problem voluntarily.

* Gantman, Alex. "Security side-effects of Word fields." Bugtraq Archive. See www.securityfocus.com/archive/1/289268/2002-09-09/2002-09-15/2. See also Lemos, Robert. "Microsoft warns of thieving Word docs." CNET News.com. September 12, 2002. See http://news.com.com/2100-1001-957786.html?tag=fd_top.

† Leonhard, Woody. *Woody's Office Watch*. September 18, 2002. See www.woodyswatch.com/office/archtemplate.asp?v7-n44.

Stuck Inside of Pages with Those Unprintable Blues Again

How typical is this scenario: You receive a 20-page Word document from a client, and you need to print it. The document contains graphics—in particular, slides copied from PowerPoint right into the Word doc. But nothing comes out of the printer. Turn the printer off, restart, and wait for the system

and printer to rethink everything . . . but still the pages don't come out. Hell freezes over, but the document won't print. Word docs are notoriously buggy when it comes to printing graphics, especially "objects" from other programs like PowerPoint, and even more especially when those objects include text.

TIP FOR WORD ADDICTS

You are nearly always better off copying the slide from PowerPoint and using Paste Special in Word to paste it as a "picture" rather than an object. What's so special about Paste Special? It gives you choices besides the lame choice offered by Word, which is to paste it as an object. You can use Paste Special to paste the slide as a bitmap image, which may get through the printer; otherwise, your printer may hang indefinitely as it tries to decipher the PowerPoint object. You would think that after two decades inhabiting the same labs in Redmond, the PowerPoint and Word development teams should be able to work together to get graphics and text to print together.

Word is supposed to be a text editor that offers WYSIWYG (What You See Is What You Get).^{*} It means, roughly, that what you see on your computer monitor is the same as what you get on the printer and vice versa. The first true WYSIWYG editor was a word processing program called Bravo. Invented by Charles Simonyi at the Xerox Palo Alto Research Center in the 1970s, it became the basis for Simonyi's work at Microsoft, including Word.

Somehow, something went awry, because now it's "You Don't Always Get What You Want"[†]—everything changes from computer to computer depending on which fonts are installed. A document produced with Word on one computer may end up with radically different formatting and pagination on another computer, even using the same version of Word.

The reason for these and other printing and formatting anomalies is this: Word silently reformats a document based on the computer's printer settings and fonts. This is bad news for certain kinds of documents, such as forms, that rely on elements precisely positioned on a page. In other words, Word documents are not guaranteed to look and print the same way on every computer and printer. The document's fonts may not be available on another computer, and the substitute fonts force the reformatting and cause pages to break in strange places. You can thank, among other things, the competing technologies for rendering fonts, which have befuddled the desktop publishing industry for two decades.

Maybe you could care less about fonts, as long as you get the document printed. But some of us want the printed document to look vaguely the same from one printer to the next. Thanks to the many differences in fonts and character spacing from one computer to the next, and from one printer driver to the next, you can't trust Word documents to look the same.

^{*} Pronounced "whiz-ee-wig." Thanks to The Dramatics for "What You See Is What You Get," released in 1971, and also to Tina Turner for "What You Get Is What You See," released in 1986; as for Britney Spears, "What U See (Is What U Get)" appears to be a rip-off.

[†] Thanks to the Rolling Stones, or rolling fonts in this case.

TANGLED UP IN THE FONT WAR

Printing problems can be traced back to the legendary Font War of the late 1980s, in which Microsoft and Apple faced off against the father of desktop publishing, Adobe Systems. As everyone knows, a digital font is a mini-program that enables a system to display and print text with a typeface (such as Palatino) set to a particular size (such as Palatino 12). While Adobe didn't exactly invent this concept, the company *did* invent PostScript and a font format that works with PostScript (the Type 1 format). The combination enabled computers to print with high-quality typefaces on different laser printers and with better quality using the same font on high-resolution imagesetters. This combination revolutionized high-quality printing in the late 1980s, and most of today's printers use PostScript.

Adobe's font format dominated desktop publishing until Microsoft and Apple—strange bedfellows at that time—developed the TrueType format to challenge Adobe's dominance. They did it to try to force Adobe into opening its proprietary Type 1 format. However, typographers weren't crazy about TrueType's quality. But even typographers have to eat, so these two formats now dominate computing with an uneasy truce. Both formats work with Windows, Macintosh, Linux, and Unix systems.

The industry giants involved in the Font War were so embarrassed by their greed that they joined together eventually to impose yet a new format, called OpenType, that slapped TrueType and Type 1 together. "That was 1996," wrote Clark Kim in *Magazine World* in 2002. "OpenType today is as popular as the U.S. Olympic hockey team in Canada."* OpenType is supposed to work with everything. It is also supposed to provide richer linguistic support and advanced typographic control. While Microsoft and Adobe support it, the gaggle of small type foundries around the world are not yet on board. The transition to OpenType hasn't been easy. It is technically challenging to do OpenType fonts, and they've already got their hands full with Type 1 and TrueType.

Word uses the fonts installed in your Windows or Mac system. When you first install Windows, only a limited number of fonts are available, but as you install other software, other fonts are added to Windows like new genes to the gene pool, and those fonts automatically become available to Word. As we all merrily computed our way into the 21st century, our systems sprouted different fonts from all these different installations. When you create a document on one system, using its fonts, and then transfer that document to a different system, different fonts are substituted, with unpredictable results.

Conversion programs exist, and fonts in both formats are ubiquitous. When they show up in your system bearing the same name (for example, Palatino in either Type 1 or TrueType formats), your system and printer can get as bewildered as you must be at this point.

* Kim, Clark. "For Font's Sake." *Magazine World*. 2002. (See <http://magazines.humberc.on.ca/magworld2002/talkingtech/font.html>.)

Adobe created the PDF (Portable Document Format) standard to enable the exchange of documents without printing problems, but Microsoft has never put much effort into making its PDF export function work well—possibly because Microsoft would rather people use the Word doc format rather than a portable format.

TIP FOR WORD ADDICTS

So you want to muck about in Word's preferences and try to alleviate the reformatting problem so that Word docs look more similar from computer to computer and printer to printer? It's not that easy. Word gets font metrics information from your output device; every time you change the print driver (even switching from one PostScript driver to another PostScript driver, or using the same driver for a different printer), text may be reformatted, pages may reflow, and line breaks may be different. Printer margins (i.e., unprintable areas defined by the output device) also influence the reformatting. Word offers an option to cut down on this reformatting: the Use Printer Metrics To Lay Out Documents switch in the Tools ▶ Options ▶ Compatibility window should be turned off. However, it doesn't eliminate the reformatting. Simply moving a document from the Windows 95/98 system to Windows NT or Windows XP can cause reformatting. Good luck with that.

Get Angry about Word Docs

There is only one way out of Word madness: stop sending Word docs to people. Only Word can open these files, so by sending Word docs, you force other people to use Word. The specification for Microsoft Word documents is a closely guarded secret, and since Microsoft chooses not to create versions for other operating systems (such as Linux), people who use those systems are left out in the cold. Keep in mind that Word doc files produced with one version of Word might not even be readable by other versions of Word—forcing others to upgrade their versions. The Word doc format is not a true standard. Microsoft changes it from time to time, most likely to force users of older versions to buy the latest version.

Richard Stallman, the founder of the Free Software Foundation, believes the use of the Word doc format hurts us as consumers and hurts the industry in general. “The worst impact of sending Word format is on people who might switch to free systems: they hesitate because they feel they must have Word available to read the Word files they receive. The practice of using the secret Word format for interchange impedes the growth of our community and the spread of freedom.”*

Whether you move off Word or continue to be addicted to it, you should know how to save your documents in a file format that others can use and that can be attached to an email without worry. Standards for document files exist, and even Word supports them to some extent.

Rich Text Format (RTF)

RTF files are readable and useful across systems and applications. RTF files preserve some font information (such as italics, bold, font sizes, and so on), and people can import the files into their word processors (including Word) to edit them. You can be sure that the RTF file doesn't harbor any viruses, because it doesn't contain any macros.

* Stallman, Richard. “We Can Put an End to Word Attachments.” GNU project. See www.gnu.org/philosophy/no-word-attachments.html.

STOP SENDING WORD DOCS!

If you must continue to use Word, at least try not to inflict this punishment on the rest of us. Free Software Foundation founder Richard Stallman encourages people who dislike receiving Word doc files to follow his example:* “For about a year, I’ve made a practice of responding to Word attachments with a polite message explaining why the practice of sending Word files is a bad thing, and asking the person to resend the material in a nonsecret format. . . . If we all do this, we will have a much larger effect. People who disregard one polite request may change their practice when they receive multiple polite requests from various people. We may be able to give ‘Don’t send Word format’ the status of netiquette, if we start systematically raising the issue with everyone who sends us Word files.”

As a result, a minor revolt is spreading across the Internet, as people post web pages explaining, in an ever-so-polite manner, why they will not accept Word attachments:

- “MS-Word is not a document exchange format” by Jeff Goldberg (www.goldmark.org/~netrants/no-word/attach.html)
- “Don’t send Microsoft Word documents to me” by Jonathan de Boyne Pollard (<http://homepages.tesco.net/~J.deBoynePollard/FGA/dont-send-word-documents.html>)
- “Please don’t send Word Documents by email” by Tobias Brox (www.cs.uit.no/~tobias/NEW/word-rejection.html)
- “Avoid e-mail attachments, especially Microsoft Word” by Neal McBurnett (<http://bcn.boulder.co.us/~neal/attachments.html>)
- “Attachments in proprietary formats considered harmful” by Manuel Chakravarty (www.cse.unsw.edu.au/~chak/email.html)

* Stallman, Richard. “We Can Put an End to Word Attachments.” GNU project. See www.gnu.org/philosophy/no-word-attachments.html.

The RTF files produced by Word *can* get crazy if they include complex formatting, tables, graphics, or objects embedded from other Microsoft applications, so you may want to simplify your document first. On the other hand, whatever application you use to open the RTF file might simplify things for you—the text will appear (more or less) the same, with page breaks, italics, tables, and possibly even footnotes, but most of the other Word-related junk not recognized by your application will be ignored.

NOTE *RTF uses the American National Standards Institute (ANSI), PC-8, Macintosh, or IBM PC character set to control the representation and formatting of a document, both on the screen and in print. See the Rich Text Format specification at <http://latex2rtf.sourceforge.net/rftspec.html>.*

TIP FOR WORD ADDICTS

As a general rule, you can lose the revision logs and other Word weirdness by saving the Word document as an RTF file that people can edit easily or as a PDF file that looks great but can’t be edited easily. Choose **File ▶ Save As** and pick the RTF file format from the Format drop-down menu. Creating a PDF file is more complicated, thanks to Microsoft—see the section “PDF Was Made for This” later in this chapter.

Portable Document Format (PDF)

PDF files preserve the look, the feel, the fonts, the graphics, the pagination, headers, footers, footnotes . . . everything. Everything is exactly as it should look.

PDF is ideal for distributing finished documents, including PowerPoint presentations and Excel spreadsheets. It's especially useful for forms that must print the same way on every printer and look the same on every computer. Even the fonts are taken care of—you can automatically embed into the PDF file the font characters the document needs. The main drawback is that people can't easily edit the text unless they buy Adobe Acrobat. Unfortunately, PDF files are very large, and PDF files from untrusted sources might also carry viruses. But the risks of catching a virus by viewing and printing a PDF file are slim. See the section "PDF Was Made for This" later in this chapter.

Plain ASCII Text

Plain ASCII (American Standard Code for Information Interchange, pronounced "ask-ee") text files contain only the text of your document with no formatting whatsoever (no fonts, no spacing, and tables are turned into text spaced with tabs). Plain ASCII files can't harbor viruses (except perhaps as source code that can't execute by itself), which makes them safe. All text editing programs can edit ASCII text files, which date back to the Model-T era of the computer industry. Any decent spreadsheet program can read an ASCII file's tab-delimited characters saved from a table and convert the mess back into a table.

NOTE *Programmers use ASCII text files to write programs, but the text files have to be interpreted or compiled into code for the programs to work.*

ASCII is a standard developed by ANSI to define how computers write and read characters. It was designed at first for teletypes and extended for displays and modern printers. If you've received ASCII text files, then you know that the text might all be in a single line, 10,000 words long, or have weird characters and control codes scattered about. It might appear stunted with only 35 characters per line, and you spend half a day positioning the cursor just so and deleting backward to remove these line breaks (an operation likely to crash Word).

NOTE *The ASCII set of 128 characters includes letters, numbers, punctuation, and control codes (such as a character that marks the end of a line). Each letter or other character is represented by a number: an uppercase A, for example, is the number 65, and a lowercase z is the number 122. (Software engineer Jim Price has a nice ASCII chart at www.jimprice.com/jim-asc.htm.)*

ASCII ART

Ever see those “art” printouts of ASCII characters arranged so that they appear, at a distance, to be a beautiful nude woman? Artists and amateurs have applied the character set to works that range from interesting emoticons (beyond the smiley :-)) face) to extreme photo-realistic digitizations of portraits. Check out ASCII Artwork (www.textfiles.com/art) for a gallery of ASCII art.

Hypertext Markup Language (HTML)

HTML files contain the codes for formatting web pages (and email). HTML is essentially ASCII text with formatting *tags* neatly separated by angle brackets from the rest of the text. Today, HTML files can contain code such as JavaScript, which hackers can use to link to other websites and embed viruses, spyware, adware, and other bad things, but plain HTML (without code) is safe. Unfortunately, Word doesn’t create a plain HTML file—it adds a lot of stuff you probably won’t want or need (see Figure 4-3).



Figure 4-3: Word puts all this junk at the top of every HTML file it creates, guaranteeing that you won’t know what the heck is going on inside. You can get rid of most of it and use the standard HTML codes for defining a header and title.

Aliens Kidnapped My Characters!

If you’ve ever created an ASCII text file from a Word doc, or opened one in Word or some other program, you probably ran into problems with Word’s special characters. What’s so special about em dashes, single and double

quotes with curls, and the sequence known as ellipses (three periods in a row)? They've been around for long time, included on typewriter keyboards since at least 1914.*

Proper style calls for using these special characters to enhance readability. But Word on the PC seems to sabotage the documents so that they can't be used with the Mac version of Word without considerable reformatting. You get these funny characters (see Figure 4-4) that look like they're part of an alien alphabet from a distant planet. You also get them when you convert Word docs to plain ASCII text.

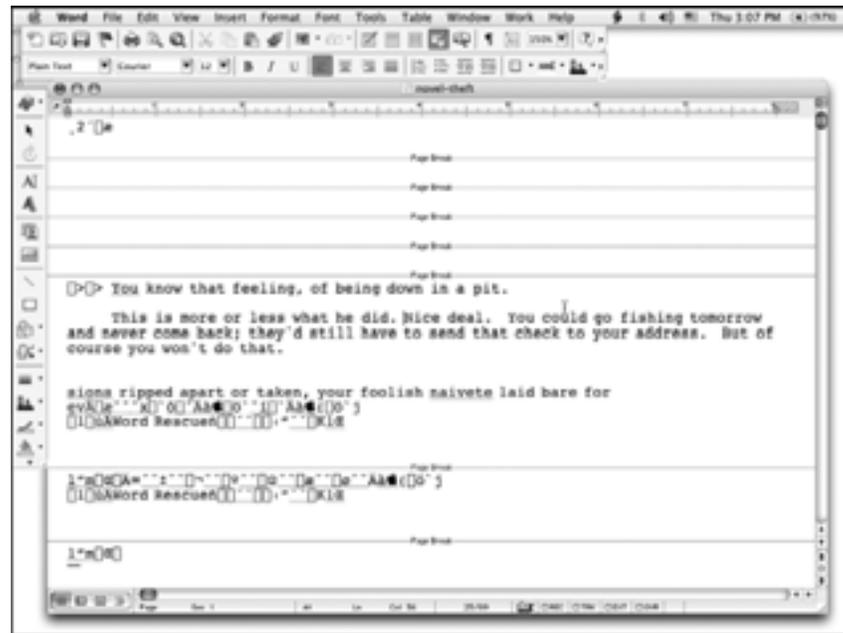


Figure 4-4: Word's special characters look like an alien alphabet when converted to ASCII text, or when converted from the PC to the Mac version of Word (or vice versa).

If you continue to use Word, you might want to consider at least converting these special characters to something else before sending ASCII or HTML versions of your document. You don't have to convert anything for PDF versions, because the special characters look fine. Even if you just use Copy and Paste to put Word text into an email message, you may find these characters lurking about in your message. Here are quick steps to change these special characters *before* converting to ASCII or HTML:

1. Choose **AutoCorrect** from Word's Tools menu. This opens Word's AutoCorrect dialog box.
2. Select the **Auto Format As You Type** tab, and uncheck the option 'Straight quotes' with 'smart quotes', as shown in Figure 4-5. (Click to remove the check mark, turning it off.)

* See the Virtual Typewriter Museum at www.typewritermuseum.org.



Figure 4-5: Don't let Word's smart quotes get in the way when converting Word docs to the ASCII or HTML formats or pasting copied Word text into an email.

But don't stop there! You still have to use the Replace command, replacing all instances of a single quote (curled) with a single quote (straight):

3. Choose **Edit** ► **Replace** and type a single quote for the Find box and the same single quote for the Replace box (yes, that's right—the same character). Click Replace All to perform the replace operation for the entire file (or selected text). Word automatically changes each curly quote to a straight quote that can pass the ASCII test.

You can replace the characters in only a portion of the document by first selecting the portion and *then* clicking **Replace All**. The replacements occur only in the selected text.

4. Repeat step 3 with double quotes. Word automatically changes the curly quotes to straight quotes that can pass the ASCII test on all computer systems.
5. Repeat step 2, unchecking the option to the replace symbol characters (–) with symbols (—). Then repeat step 3 with em dashes (replace with a double dash, as in “—”), en dashes (replace with a single dash), and ellipses (replace with three periods in a row).

Use Alternatives to Word

When you're a hammer, every problem looks like a nail. (And when you're a nail, you wish there were no hammers around.) And when you're a Word user, every problem looks like it needs something from Word's menus or toolbars to solve.

The industry needs an alternative, if only to keep Word honest. Word is expensive (more than \$200 for the stand-alone retail version, as this is written), tends to be a memory hog, and the built-in macro features are inherently insecure, providing an excellent breeding ground for viruses. Documents generated by Word also have the bad habit of hiding extra information (such as deleted text and personally identifiable details about the file author) in their headers. And files created in Word usually aren't readable by other word processing software, because Microsoft keeps Word's document format proprietary.

If you use Word on a PC, OpenOffice.org Writer offers the easiest transition to an alternative. WordPerfect also offers a look and feel that comes close to matching the Word's menus and toolbars, but OpenOffice.org Writer (hereafter known as OOo Writer) matches Word on the PC function for function and even looks just like it. If you use Word on a Mac, you can also use OOo Writer, but it looks like the PC version, which is somewhat of a downer. But the OpenOffice.org volunteers have scheduled a native Mac OS X version to come out real soon now. I used OOo Writer on the Mac in its current form to write this book.

You can fool people into thinking you are using Word by using one of these alternatives and then saving your documents in the Word doc format—just for them. For everyone else, you can save RTF documents that others can open and edit in any word processing program, or save PDF files that can be opened by anyone using Adobe Reader or similar PDF viewer. These alternatives also have their own native file formats, which are not as ubiquitous as the Word doc format and therefore not as prone to carry viruses.

WordPerfect: The Legal Remedy

For a commercial alternative that costs about the same or less than Word, consider WordPerfect. Lawyers and legal assistants swear by WordPerfect for legal contracts and briefs, possibly because it has been around long enough to be entrenched in that professional industry. It was the first word processor to offer automatic paragraph numbering and automatic footnote numbering and placement on pages. WordPerfect used almost every possible combination of function keys with the CTRL, ALT, and SHIFT keys, making it easier to use than the word processor of the day (WordStar, my first love).

Originally written for Data General minicomputers, WordPerfect migrated to the IBM PC in 1982. The program's popularity took off with version 4.2 in 1986, and in 1989, WordPerfect 5.1 for DOS became the word processing market leader. Then Microsoft moved in for the kill.

Microsoft started with WordPerfect's function keys, pre-empting them by creating incompatible keyboard shortcuts for Windows 3.0 that conflicted with them (e.g., ALT-F4 became Exit Program instead of WordPerfect's Block Text). The DOS version's impressive arsenal of finely tuned printer drivers was also rendered obsolete by Windows' use of its own printer device drivers.

“Microsoft knows that the technologically perfect product is rarely the same as the winning product,” wrote James Gleick in *The New York Times Magazine*. “Time and again its strategy has been to enter a market fast with an inferior product to establish a foothold, create a standard, and grab market share.”*

One problem was that WordPerfect’s function-key–centered user interface did not adapt well to the Windows paradigm of mice and pull-down menus. Another problem was that WordPerfect took too long to migrate to Windows 3.0, which became the most popular operating system ever. But the main problem was that Microsoft engineers had advance information about Windows.

Novell, the company that bought and later sold WordPerfect in the mid-1990s, eventually sued Microsoft for restraint of trade, charging that Microsoft withheld “critical technical information” about Windows. Novell also charged that Microsoft deliberately excluded WordPerfect from the marketplace by using monopoly power to prevent hardware manufacturers from offering WordPerfect to customers.†

Microsoft defended itself by pointing out that its application development group (responsible for Word) and its system group (responsible for Windows) were simply able to “fly in formation,” as CEO Steve Ballmer liked to put it. Only this formation was more like a bomber run.

WordPerfect eventually landed at Corel as a much better, cleaner, and faster product with some excellent features that Word lacks. For example, it shows a preview, right in your document, when you are considering changing the font, size, or alignment of the text, allowing you to change your mind before committing to alterations. It offers a “reveal codes” feature (see Figure 4-6) that lets you “open the hood” on any section of text and make changes to formatting by changing the codes directly. As a result, WordPerfect is less likely than Word to surprise you with automated reformatting. It also uses its own settings for print options rather than forcing you to set them each time in the Windows print driver.

WordPerfect 12 Office Suite includes Quattro Pro for spreadsheets and Presentations for presentations. It can import and save various Word and Office doc formats and RTF files with almost perfect back-and-forth compatibility. It can also save documents in the PDF format, with hyperlinks if you use them. WordPerfect’s “compatibility toolbars” let you save documents in Microsoft’s formats with a single click—you can even substitute icons and menus that approximate Microsoft’s user interface instead of WordPerfect’s. A Legal toolbar provides easy, one-click access to specialty tools for the legal community, including the Pleading Wizard, the Clipbook, the Concordance tool, and support for EDGAR electronic document filing. WordPerfect 12 Office is not just for lawyers; it’s a solid product for all kinds of word processing, and Corel (www.corel.com) provides adequate support that is as good as if not better than Microsoft’s.

* Gleick, James. “Making Microsoft Safe for Capitalism.” *The New York Times Magazine*. November 5, 1995.

† LaMonica, Martin. “Novell sues Microsoft for sinking WordPerfect.” CNET News.com, November 12, 2004.



Figure 4-6: WordPerfect reveals the formatting codes underneath the text for quick adjustments.

OpenOffice.org Writer: A Free Alternative to Word

OpenOffice.org (OOo) Writer is part of the open source OpenOffice.org package, which started its life as StarOffice, a competitor to Microsoft Office for IBM OS/2 and Unix systems. Sun Microsystems bought it and converted it to an open source project called OpenOffice.org. It includes Writer for word processing, Calc for spreadsheets, Impress for presentations, Draw for drawing graphics, and conversion utilities for various file formats, including all Microsoft formats.

NOTE *OOo is not some hobbyist's dream from the days of teletypes. In the first month after OpenOffice.org was released (May 2002), the project reported nearly a million downloads. Nobody knows exactly how many people use it, but most estimates put it at more than 10 million.*

OOo is free—you can download versions for Linux, Unix, Solaris, Windows, or Macintosh (OS X) from the OpenOffice.org site (www.openoffice.org). You can also get it on CD or as part of a book-CD package, such as *OpenOffice.org Writer* by Jean Hollis Weber (O'Reilly Community Press). And

here's something you can't do with Word or the rest of the Office package, but you can do with OOO: you can legally install it on as many computers as you want. You can even give it to your friends.

NOTE *If you or your company prefers a traditional relationship with a software vendor, you can purchase StarOffice from Sun Microsystems for \$75.95 for a single copy or \$50 per copy for 150. StarOffice is an enhanced version of OpenOffice.org, with additional translation filters, fonts and clip art, a manual, and a database (see www.sun.com/software/staroffice/6.0/index.html).*

A frequent criticism of open source software is that you have to pay for support, but you can get free support for OOO if you subscribe to its free mailing lists. This support is as detailed and as accurate as any you can get from most paid technical support services.

If you know Word, you can be comfortable with OOO Writer in just minutes (see Figure 4-7). OOO Writer matches Word feature for feature and even beats Word in some areas. For example, the Stylist floating palette for setting formatting styles is more convenient than Word's style catalog and extends to more than just paragraphs and characters. OOO Writer also offers text frames and lets you use styles with them, making it far easier than Word to set up multiple columns or a newsletter layout.

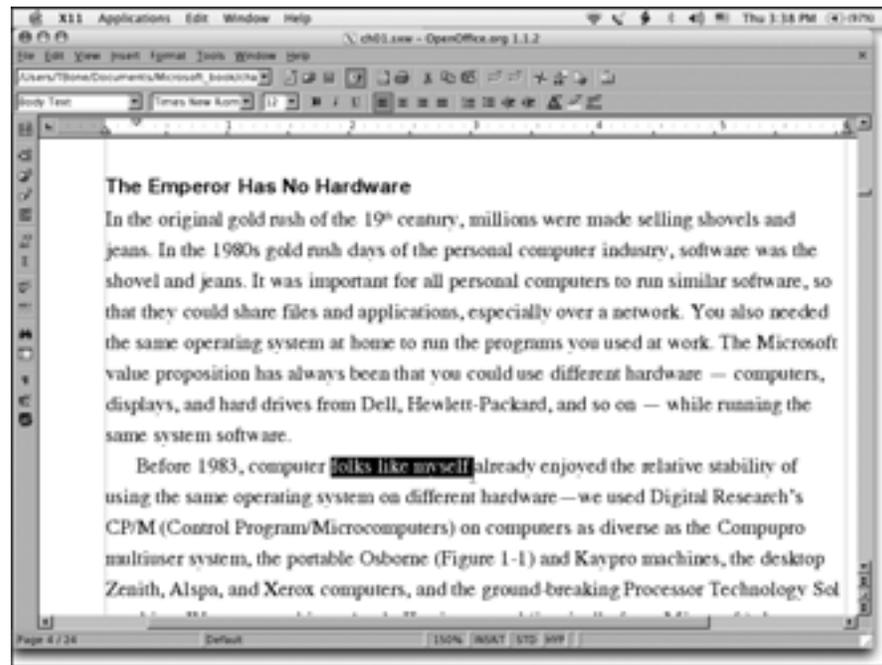


Figure 4-7: OpenOffice.org Writer looks just like Word and matches it feature for feature.

I've spent many frustrating hours with Word trying to make different headers and footers work properly in a long document. But you can control headers and footers far more easily in OOO Writer, because these elements are text

frames attached to page styles that you control from page to page. And if you've spent a night wrestling with why Word's autonumbered lists are so screwed up, you'll be happy to know that OOo Writer's autonumbering works perfectly every time.

OOo Writer's menus look just like Word's, even to the point of offering the confusing choice of Configure or Options in the Tools menu. Autoformat and Autocorrect? They're in there, and you can customize them and turn them on or off. Envelopes and labels, fields, outlining, track changes, versioning? They're in there—maybe by different names and in different menus, but they're in there. The developers of OOo were determined to provide everything Word offers, and since there haven't been many real innovations added to Word since about 1995, it's an easy target.

In all this good news, there must be some bad. Indeed, OOo Writer has trouble importing Word docs that use some of Word's more esoteric features, such as the STYLEREF field (which OOo Writer converts to ordinary text). I've also discovered anomalies importing Word docs that use cross-references—although you can use OOo Writer to create cross-references without a problem. Forms don't translate well, and macros are either ignored or preserved for future use with Word (if you intend to save your document in the Word doc format).

TIP *OOo Writer imports Word files without translating or executing macros. The program does not support Visual Basic or any form of internal macro, so even if the document's macros contained a virus, it most likely would not harm your computer. The program does give you the option of preserving macros in Word documents, so that they can be used when someone opens the document in Word again. In other words, you can import the Word doc into OpenOffice.org Writer, bypassing any virus, save the document back in Word format, and pass that virus along to someone else who still uses Word. But it wouldn't be very nice to do that.*

OOo Writer can save documents in various Word doc formats—including Word 6.0, 95, and 97/2000/XP formats, as shown in Figure 4-8. Assuming the Word user has the same fonts you used, the document should look the same, except that the bullets in a bulleted list may use an unexpected character rather than a bullet symbol. (You can avoid that problem by setting the bullet character in OOo Writer's Options tab for the list style.) Without the same fonts, the document might have differences in pagination, paragraph length, and paragraph alignment. You can also import or save HTML or RTF files and save a PDF file that looks exactly like the document in OOo Writer, fonts and everything, no matter what computer you use to view and print it.

Considering the difficulties of writing translation filters for importing, exporting, and saving files in different formats, what is unusual is not that OpenOffice.org's filters have problems, but that they have so few. Still, you should be prepared to do a bit of manual cleanup when importing Word docs into OOo Writer (or into any other program for that matter).



Figure 4-8: OpenOffice.org Writer can import and save various Word doc formats as well as RTF.

AbiWord: A Simpler Free Alternative to Word

AbiWord is part of a larger project known as AbiSource (www.abisource.com) started by the SourceGear Corporation, which later released the source code to a developer community that quickly formed around the project. The project's goal was to come up with a free office suite that would run on any operating system out there, and the community has largely succeeded, with more than 2.5 million downloads. AbiWord runs on Windows, Linux, Mac OS X, FreeBSD, and Solaris. It supports right-to-left, left-to-right, and mixed-mode text for European, Hebrew, and Arabic languages, and offers dictionaries for more than 30 languages.

AbiWord looks just like Word (no surprise there), except its icons are larger (see Figure 4-9), it loads faster, and it has less of a memory footprint, using only about 6 MB of RAM compared to Word's 30 MB or more. It doesn't try to match Word feature for feature, but it offers most of the features you'd expect from a word processor. You can employ tables, bullets, lists, images, footnotes, endnotes, and styles, just like Word. It even offers mail-merging for creating form letters and replacing fields with text from databases or spreadsheets.

If you are familiar with Word, you can be very comfortable with AbiWord; if you have no experience with Word, you won't miss all the obscure features that make Word a bloated memory hog. You may prefer AbiWord over OpenOffice.org Writer for the same reason—simplicity. AbiWord's code is written

to be tight and fit in small memory spaces; to do this well, it separates some of the extraneous functions into plug-ins. AbiWord lets you install plug-ins such as a dictionary, a thesaurus, a Google searcher, and translators from Babelfish and FreeTranslation. Also available are AbiPaint for painting simple images and graphics, and AbiGimp for photo and image manipulation. The Google plug-in lets you search Google with your selected text.

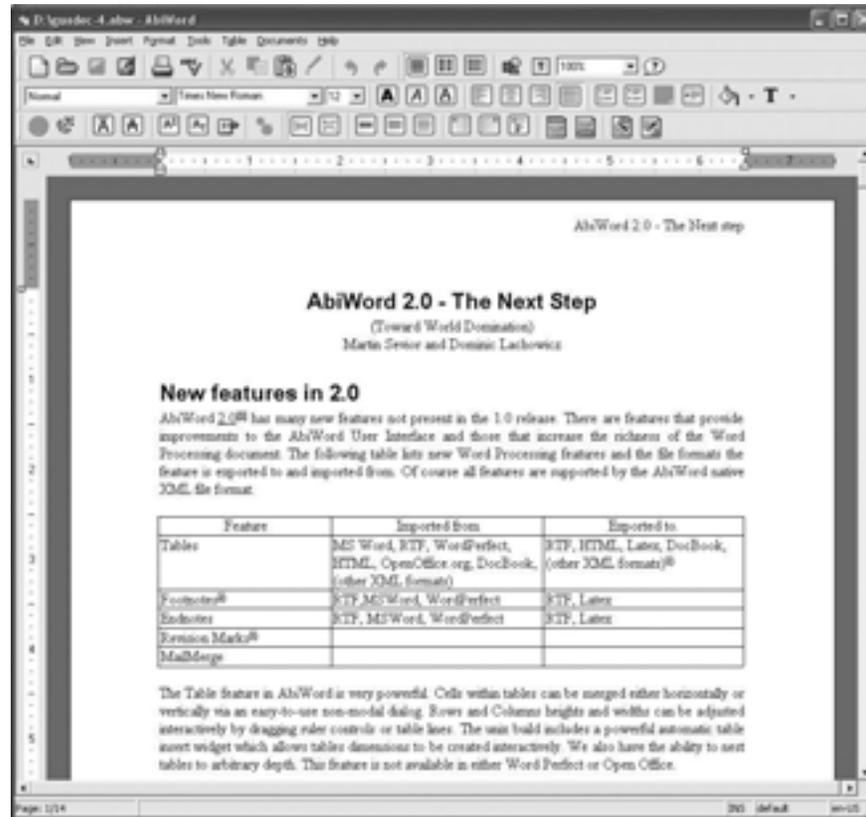


Figure 4-9: AbiWord, a free open source application, looks just like Word but uses less memory and runs on just about every type of computer system.

AbiWord can import Word docs, and import and export OpenOffice.org and WordPerfect docs, as well as RTF, HTML, and even XHTML files. It even offers a command line for executing some of its functions, to make it easier to run operations on a server, such as generating form letters, printing documents, or converting documents to another file format. Like OpenOffice.org, AbiWord saves a document in a compressed ASCII text file with XML (eXtended Markup Language) markups, which makes the document file easy to open and read using any text editor and easy to process with XML tools. While OpenOffice.org Writer gets more publicity as an open source alternative to Word, AbiWord is a lean, fast, and cross-platform open source alternative that looks more like Word, especially on a Mac.

How to Collaborate on Documents without Word

Let's get down to brass tacks: How do you collaborate with all those other people out there who are stuck using Word?

Here's what I do: These days I use a Mac PowerBook running OS X, but I could be using a PC laptop just as easily, running Linux. Either way, I can use a free text editor, such as TextEdit on the Mac, to write all my text—book chapters, white papers for high-tech companies, web pages, and email messages I want to compose at my leisure before sending.

I then copy the text into OOo Writer, and I apply styles for book chapters and whatnot, controlling the page layout. I can send the finished draft in RTF or Word doc format. The recipient can't tell that I used OOo Writer.

Collaboration is a tougher nut. Most people assume that means allowing other people to make changes directly to your document. This is the Word model—send out a Word draft with Track Changes turned on, so that people can make changes and see other folks' changes. People can also add comments that others can see. I can use OOo Writer to import a Word doc with tracked changes, add my own changes, and save them back to another Word doc, but I don't like doing it.

The problem is that you have to trust everyone. People could turn Track Changes off, make changes, and then turn it back on, and no one would be the wiser (unless they laboriously compared the text from the one version to the next). Even if you trust everyone, if you forget to turn Track Changes on for reviewers, they might make changes that aren't tracked without realizing how much of a problem they may be causing. Heck, in Word, you can't even track some of the changes made to tables (fortunately, it warns you). There is no secure way to collaborate using the Word doc model.

To collaborate securely—by this I mean make drafts available for review comments and incorporate them back into the draft—I export a PDF version from OOo Writer and invite comments. When I receive commented PDFs back from reviewers, I make the changes myself, in OOo Writer, based on the comments. This is a far more secure method of collaboration. It doesn't require retyping—you can copy text from a PDF comment and paste it right into a document without any trouble. (The next section provides more details about using PDF.)

There's another way to collaborate that lets others make changes directly—it's just not as secure, and you have to trust the reviewers. I submit my book chapters this way. I send the first draft in the OOo Writer format to my editors, who make changes and add comments (OOo Writer offers tracked changes and commenting just like Word) and send them back to me. You can even use OOo Writer to import Word docs with tracked changes and comments and save a document back in the Word doc format with the same tracked changes and comments (and your own ones added in).

Collaboration with documents can quickly get out of hand when others don't have the same fonts you have. Character sizes and spacing can change everything from the way lines break to the pagination and placement of footnotes. You might be referring to page 10 while someone else opening the

document on a different system with different fonts would need to look at, say, page 9 or 11. And you can forget about trying to collaborate on complicated page layouts using Word, OOo Writer, WordPerfect, or any other doc file format. PDF was made for this.

PDF Was Made for This

I was there when John Warnock, co-founder of Adobe Systems, first unveiled PDF to analysts in 1991. With a gleam in his eye and a silent nod to the analysts who had written extensively about the Font War that had nearly been his undoing, Warnock introduced PDF as if it were already a standard and read a statement that has since become the PDF manifesto: “Imagine being able to send full text and graphics documents—this means newspapers, magazine articles, technical manuals, and so on—over electronic mail distribution networks. These documents could be viewed on any machine, and any selected document could be printed locally. This capability would truly change the way information is managed.”*

Okay, I thought, sounds wonderful so far. PDF would solve many of the print problems plaguing the desktop publishing industry. It would do so by encapsulating the font, graphics, and page layout information in a standard format everyone could use. Indeed, by 1996, PDF had become a standard in the high-quality pre-press and color printing industry. But why not provide a format that people could *edit*? I asked Warnock this question back in 1991, and he pointed out that the leading page layout programs, namely Aldus PageMaker and QuarkXPress, used proprietary formats, and these companies would never agree to cooperate at the editable document level. If they had agreed back in 1991, the Word doc format would not be so ubiquitous. By 1992, I was using PageMaker to compose a newsletter, sending it to the printer as PDF and exporting it to the Word doc format whenever I needed to spin off an article for a magazine that required Word doc submissions for additional editing.

The Adobe PDF file has since become the de facto standard for distributing documents in a secure, reliable way. The lack of an editing capability is, in this scenario, a blessing, because you can distribute a PDF and know that the text can't be altered or copied without your knowledge—you can even digitally sign the PDF so that readers can verify its authenticity. With a track record of more than a decade, PDF has been adopted by governments and enterprises around the world to reduce reliance on paper. You already use it today to file your U.S. income taxes. It's the standard format for the electronic submission of drug approvals to the U.S. Food and Drug Administration (FDA) and for electronic case filing in U.S. federal courts. It is also the standard format used for advertisements in newspapers and magazines.

Why is PDF so good for distribution? It preserves the fonts, images, graphics, and layout of any source document, regardless of the application used to create it and regardless of what fonts you have in your system. PDF

* Warnock, John. “John Warnock's ‘Camelot’ signalled birth of PDF.” Planet PDF. January 18, 2002. See www.planetpdf.com/enterprise/article.asp?ContentID=6519.

files can be shared, viewed, or printed by anyone using the freely distributed Adobe Reader software (which you can grab from www.adobe.com; more than 500 million copies have already been downloaded). PDFs include embedded fonts and color profile information for more accurate color rendering across different systems. You can expect the PDFs you create to look the same way on another computer as it does on your computer.

All the word processors described in this chapter can export PDF files, even Word. In some systems, all you need to do is pretend to print the document and choose a PostScript printer driver to save the output as a PostScript file. Adobe offers the rather expensive Adobe Distiller product to convert PostScript to PDF, but there are also free applications, such as Free PDF Converter (www.primopdf.com) and Pdf995 (www.pdf995.com), and inexpensive commercial products, such as CutePDF (www.cutepdf.com). Adobe also offers server products for creating PDFs in an enterprise, eliminating the need for PDF creation applications on desktops.

Adobe never pursued the ability to edit PDFs except with its own applications. Adobe offers Acrobat for creating, combining, and exchanging PDF files, which lets you collect documents, emails, graphics, spreadsheets, and other attachments into a PDF file that preserves the integrity of the layout, whether the pages are oriented horizontally, vertically, or mixed. Acrobat also allows you to password-protect PDF files to prevent unauthorized viewing and altering, while also enabling authorized reviewers to use commenting and editing tools.

Adobe Acrobat provides excellent commenting and markup tools. You can draw attention to something with lines and arrows, draw boxes and free-form shapes around portions of text, add comments in separate windows, and even add voice comments and attachments, as shown in Figure 4-10. When you receive a PDF file with comments appended to it, you decide which comments to incorporate—and it's easy to copy and paste comments from the Adobe Reader or Acrobat window into your document. You can even use Acrobat to export the text in PDF files—including comments—into RTF files or Word docs. And you can search PDF files for words appearing in the text and in annotations, bookmarks, and data fields.

NOTE *Adobe offers an online service for creating PDFs from your documents. It not only accepts Word and WordPerfect documents, but also Adobe Illustrator, InDesign, FrameMaker, PageMaker, and Photoshop files. Visit <https://createpdf.adobe.com> to learn more.*

PDF is the easiest way out of the conundrum of using flaky Word docs and other Office docs to collaborate on documents or publish them to the Web. It is also the most useful format for sending documents to printers. The free Adobe Reader application can open, display, and print PDF files on all versions of Windows and Macintosh systems and Linux. The same PDF file will print on a cheap inkjet printer as well as on an expensive imagesetter. PDF files can contain multimedia elements like movies or sound as well as hypertext elements like bookmarks, links to email addresses or web pages, and thumbnail views of pages.

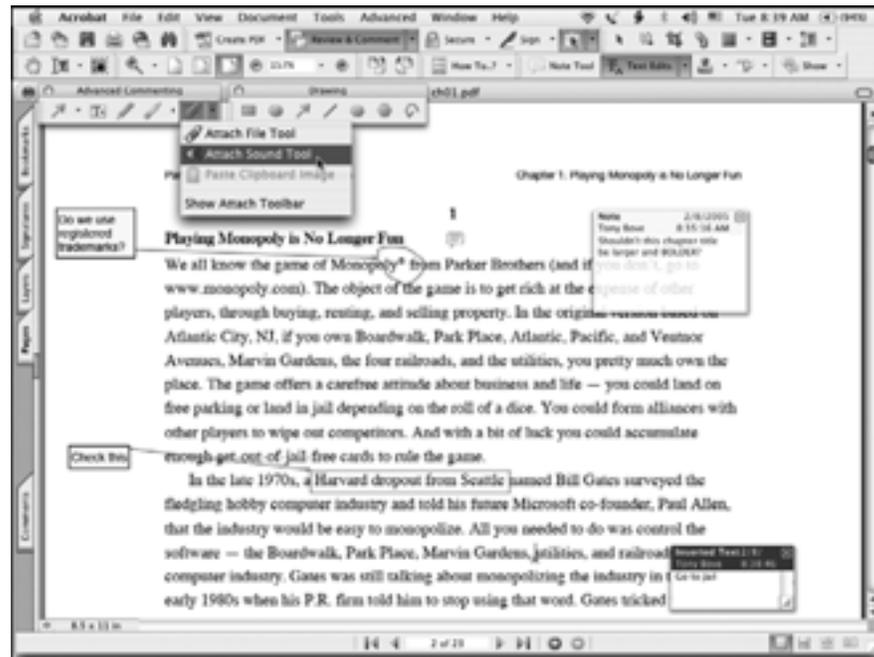


Figure 4-10: Adobe Acrobat lets reviewers mark up text for corrections, draw attention to details, add comments, and even voice opinions as well as attach files (such as background material). None of this stuff changes your original document, so you retain control.

TIP FOR WORD ADDICTS

PDF Writer is a printer driver that installs automatically into the Microsoft Office environment on Windows as a virtual printer, but don't use it. PDF Writer is just not as good as Adobe Distiller for creating PDFs with bookmarks for easy viewing or for pre-press systems and high-quality color printing. Always print to a PostScript file and then use Distiller or a similar tool to create a PDF. If you install Acrobat on a system containing Office, it adds macros to Word, Excel, and PowerPoint that automate the creation of PDF files. Whenever you select the **Export to PDF** option, the application will create a PostScript file and launch Distiller to automatically convert that code to a PDF.

Mac OS X uses PDF technology within the operating system itself. Apple calls this technology Quartz, and it runs as a layer on top of the kernel of the system, rendering all two-dimensional objects. As a result, PDF is now the standard document format for the Mac, and native Mac applications can create and import PDFs without the need for outside programs. PDF is not exactly open source, but it *is* an open file format specification, and Adobe makes information available to anyone who wants to develop tools to create, view, or manipulate PDF documents (which is how Apple developed Quartz).

With so much knowledge about PDF available in the developer communities, one wonders why Microsoft's support for PDF is so minimal. "Microsoft is fighting to keep Office as the standard archival format for documents," said Rob Helm, an analyst with Directions on Microsoft.* "If companies were to standardize on PDF, Office would become just one PDF authoring tool among many. It's a very long-term potential threat, but Microsoft can afford to look several steps ahead."

And so should you. For a look into the future, see how easy it is to display and print documents with Macs and Linux systems. They use PDF.

* Becker, David. "What's behind Microsoft's Office moves?" CNET News.com. December 1, 2003. See http://news.zdnet.com/2100-3513_22-5111644.html.