

## DocuComp — Comparing Documents the Modern Way

by Greg Guerin

So you've written that great new proposal, and sent it to your colleague in Chicago. Now it's been returned with "just a few revisions." But the proposal is 37 pages, and you haven't a clue as to where those changes might be. On top of that, you sent out the original in Word 4.0 format, and now it's come back in WordPerfect format.

And then there was that time you sent a co-worker a monster budgeting spreadsheet and it came back with "just a few numbers tweaked." No mere mortal had the patience to print out those 14 pages of dense numbers and wade through them all to find what changed. Besides, the total budget hadn't changed by even a single penny.

If this sounds like the kind of agony that computers are supposed to *eliminate* instead of produce, then read on.

**DocuComp 1.51**, from Advanced Software, is a document comparison program that could easily handle both these situations, and many more besides. It compares the text of any two documents, marks the differences, and interactively shows you both documents in a side-by-side display. It can also produce a "composite" document that shows the entire content of both documents, with additions, deletions, and intra-document moves shown in different styles, as well as a revision list of all changes.

### First Impressions

**The License Agreement** — You've seen these printed with every piece of software you buy. Sometimes you even read them. They usually say that the software company can sue your pants off if you copy the software, but doesn't promise that the software itself will do anything at all. Well, not this time. The makers of DocuComp offer a 30-day money-back guarantee, less shipping charges. This seems to be one of the more generous agreements around, but it does have one restriction: you have to write and tell them why you didn't like the software. It doesn't say they have to approve of your reason, just that you tell them.

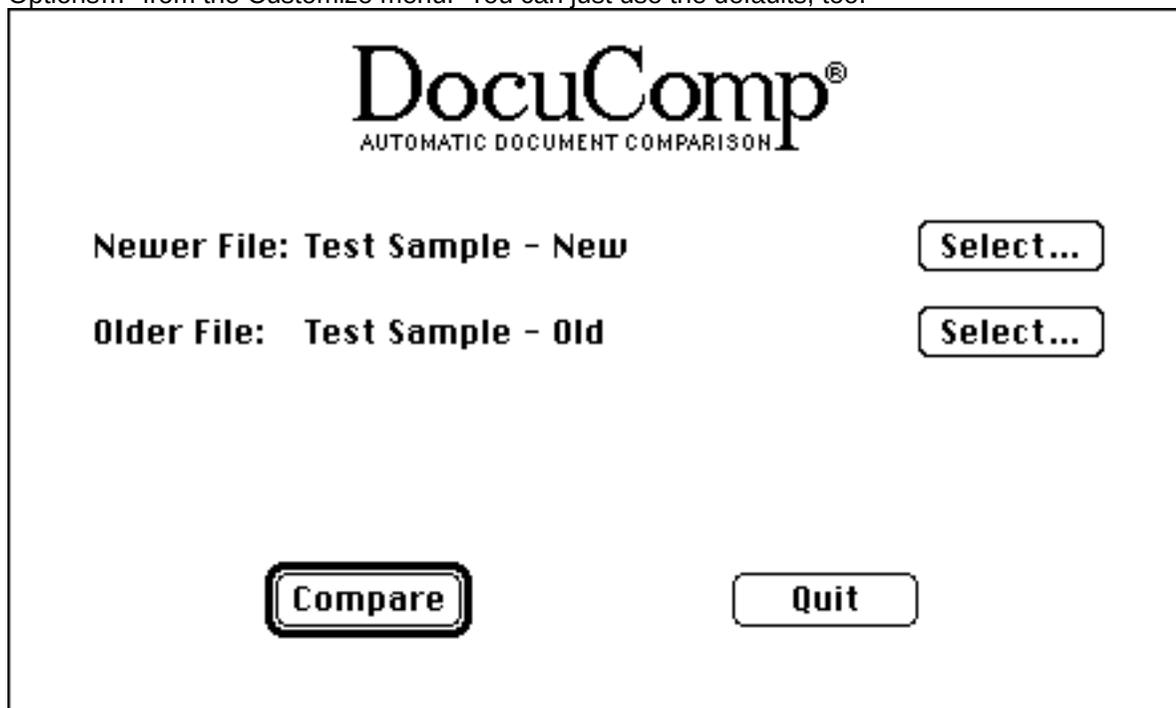
**The "Read Me First" file** — OK, so I didn't read the manual before I started, but I did read this. I usually do, because if there's a "feature" that might erase my hard disk because of some incompatibility, I want to know about it *FIRST*. No problems were listed, just a synopsis of which document formats DocuComp reads and writes, and some short tips about each of the file formats.

**The "Cold Start" Test** — I like to give most software I get the "Cold Start" test. That is, I don't read the manual at all, but just copy the program to my hard disk and run it. If it's well designed, then I should be able to easily figure out how to make it work, just relying on my previous experience with Mac software. After all, how many times do you need "Open" in the "File" menu explained to you? About the most I'll allow for "initial training" is reading the back of the package that the software came in.. Then it's time to try it out.

DocuComp comes with two example documents for comparison. I didn't use these (why make things easy?), but instead compared two revisions of a technical analysis I'd written a couple of months earlier. The result? It worked perfectly. Text I had removed appeared with a line through it, and text I'd inserted appeared in bold. It found every change I'd made, and displayed them in a way that made the changes obvious.

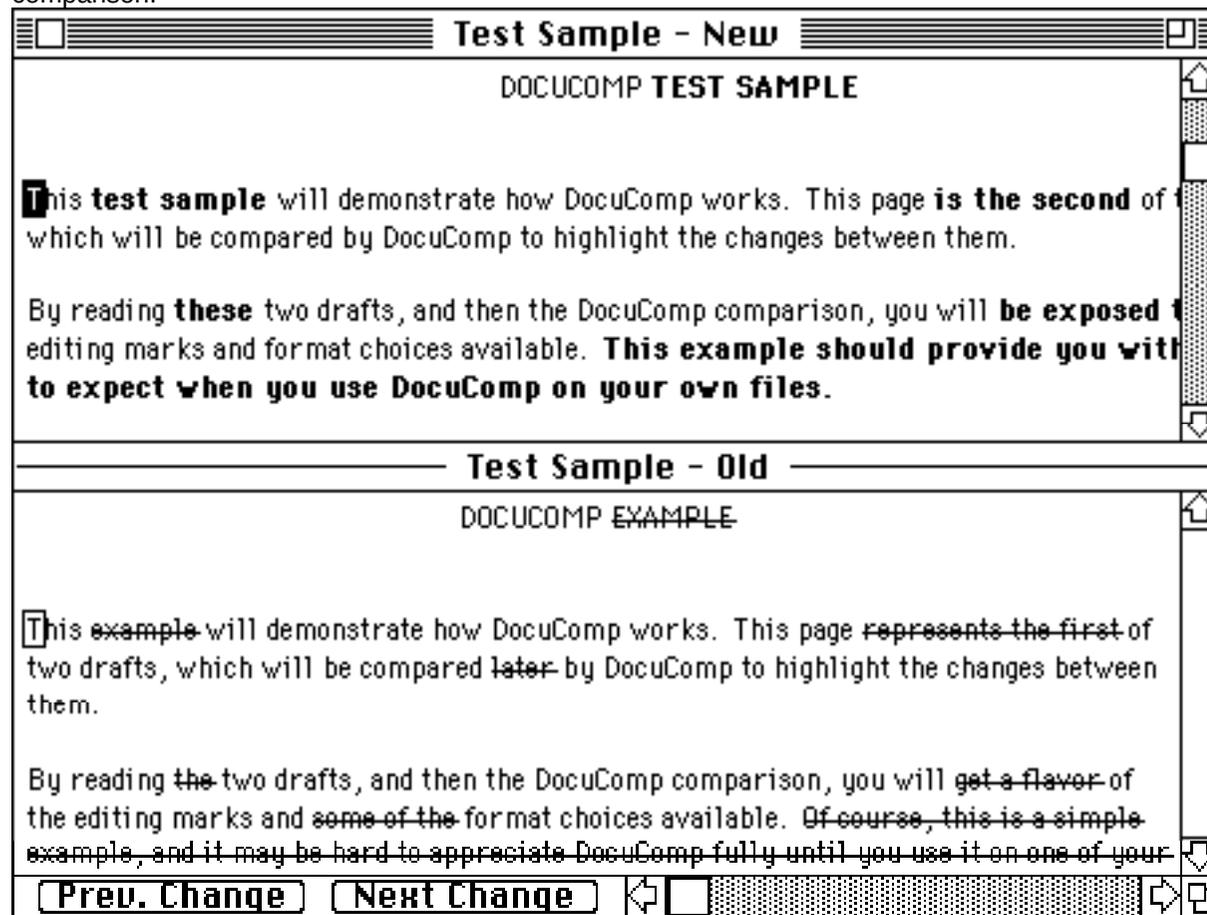
### Seeing Changes

The first window to appear in DocuComp is the "control-center" shown in Figure 1. It lets you select which files to compare. These file-names are remembered even after DocuComp quits. Before you perform the comparison, you might also want to set the styles of how changes are displayed. These changes are made by choosing "Reading Options..." from the Customize menu. You can just use the defaults, too.



**Figure 1: The Control-Center Window**

When you've selected two documents and clicked the "Compare" button, the comparison proceeds and you see the comparison window (Figure 2). This is a two-paned window, with synchronized scrolling. The newer document is displayed in the upper half, the older one in the lower half. The size of the window is determined by the size of your display, and it remembers the window position and size between invocations of the program. Synchronized scrolling means that when you scroll in one pane, the other matches it, so you are always seeing the equivalent parts of the two documents. I found this to be a very useful feature, and it seemed the most natural way of interactively viewing the comparison.



**Figure 2: The Comparison Window**

As Figure 2 illustrates, each change appears in a specific style. You can customize these, as I did, or use the defaults. Here, I've set inserted text in **bold**, deleted text in ~~strikethrough~~, and moved text in *italics*. This screen shot shows additions and deletions, but no moves. You can also set the font and size in which the comparison document will appear — this is 10-point Geneva.

The comparison window lets you scroll through the document using the scroll bars at the right side, or you can skip through from change to change using the buttons at the lower left. You can also use the arrow keys, but the manufacturer acknowledges that there is a bug in this that makes the displayed text go wacky. This is the only real problem I encountered, and Advanced Software is working on fixing it.

In addition to this interactive viewing, you can save the comparison as a file which can be read by any word processor supporting the RTF format. Word & MacWrite II are two programs that I know support this format. Others may also do so — check your manual. The file is a "composite document" containing the combined additions, deletions, and moves of both documents. DocuComp provides an array of options for saving this document. You can choose to not show deletions, or to show them only as a single character (useful in spreadsheets), and a variety of other options as shown in the "Composite Layout" dialog box in Figure 3.

Layout for Composite Document	
<b>Include:</b>	<input checked="" type="checkbox"/> Composite <input checked="" type="checkbox"/> Notes <input type="checkbox"/> Revision list
<b>Line numbers:</b>	<input type="checkbox"/> None <input type="checkbox"/> Every line <input checked="" type="checkbox"/> Every 5th line <input type="checkbox"/> Every 10th line <input checked="" type="checkbox"/> Restart on each page
<b>Page numbers:</b>	<input type="checkbox"/> <input type="checkbox"/> - <input type="checkbox"/> - <input checked="" type="checkbox"/> <input type="checkbox"/> -
<b>Line spacing:</b>	<input checked="" type="checkbox"/> Single <input type="checkbox"/> 1-1/2 <input type="checkbox"/> Double
<b>Deleted text:</b>	<input checked="" type="checkbox"/> Show <input type="checkbox"/> Omit <input type="checkbox"/> Replace with: <input type="text" value="↑"/>
<b>Revision bars:</b>	<input type="checkbox"/> None <input type="checkbox"/> Left <input checked="" type="checkbox"/> Right <input type="checkbox"/> Alternating
<input type="button" value="Set to defaults"/> <input type="button" value="Cancel"/> <input type="button" value="OK"/>	

Figure 3: The Composite Layout Dialog

Two features in Composite Layout deserve special mention: *line numbers* and *revision bars*. Having line numbers makes it much easier to review the composite document among several people. You can print it out and distribute it, and everyone can refer to "page X, line Y." This is also a boon for legal documents, which are customarily printed with line numbers.

Revision bars are vertical lines at the side of a page which mark out those areas that have changed since a previous revision of some document. Apple itself has recently been using this method in distributing its Macintosh Technical Notes, and it is a real time-saver to instantly see what's changed on each page.

When you print or save the composite document, you can hilite the changes in styles that are different from the styles displayed on the screen (Figure 4). This may be useful when you use colors on the screen, but don't have a color printer. Or you might have other reasons for using two different style sets. If you want the simple choice, there's the one-button method: "Match Screen" sets the composite styles to match the screen styles.

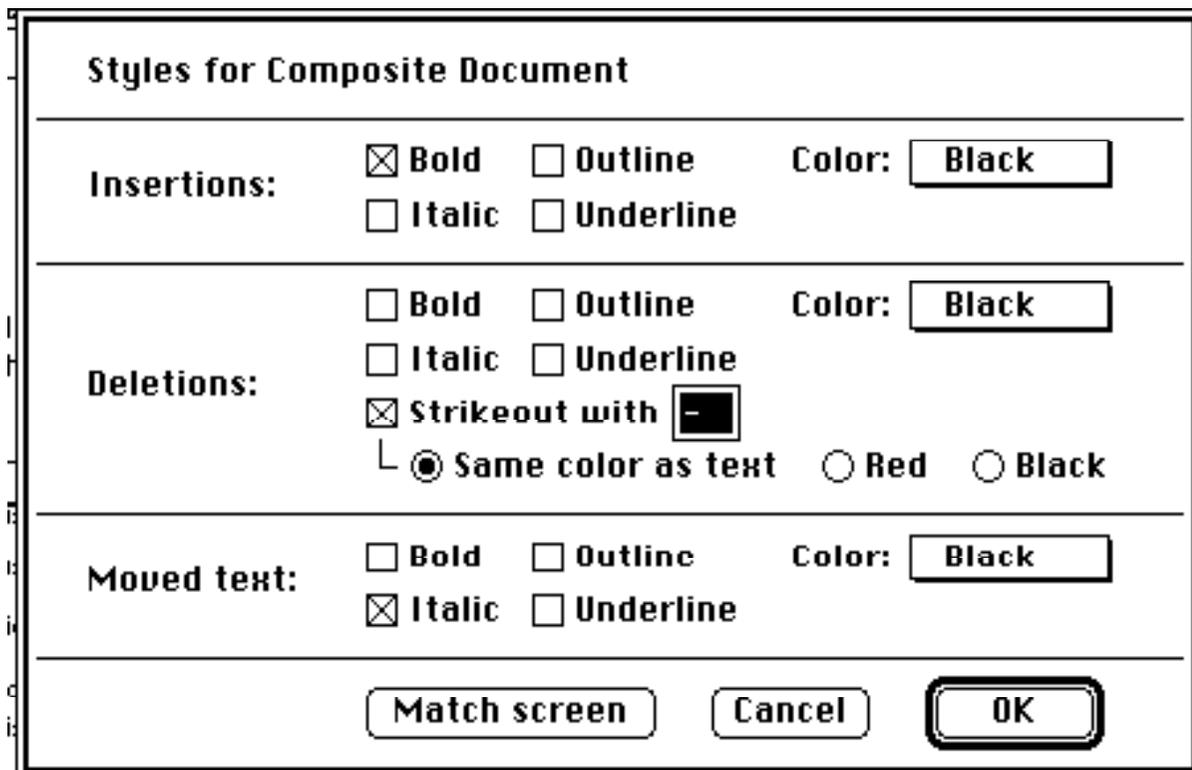


Figure 4: The Composite Styles Dialog

If you wanted a printed version of the composite document, without saving the file and opening it in your word processor, you could just have DocuComp print it. The same options for saving the composite document apply to printing it. You can choose from numbered lines, page-number placement, revision bars, etc. ... And don't forget the highlighting styles. Thankfully, there's the simple path, too. You can always click the "Set to defaults" button and get something reasonable.

With all these methods for finding changes between two documents, one last feature of the composite document layout is a true time saver: a *revision list*. When you print or save the composite document, a list of page numbers and line numbers for *each change* in the document is appended to the end of the composite document. It even shows what the change was, or the first line of it in the case of multi-line changes. So if you've got a long document and only a few changes, it pays to first look at the revision list. And if you don't want it, you can tell DocuComp not to produce the revision list.

### What DocuComp Can Read

Since not everyone has the same word processing program, DocuComp has taken the pain out of comparing files saved in different formats. It can read any of the following formats: Word (3.x and 4.x), WordPerfect (Mac or PC), WriteNow (1.x and 2.x), MacWrite II (1.x), MORE II and MORE 3, or any document saved as text-only. Unfortunately, it does not directly support the original MacWrite format. The way to work around this is to use your word processor to read the MacWrite file (most do) and then save in a form that DocuComp can read. Although this works, it's clumsy. Since the original MacWrite format is a de-facto standard on the Mac, it seems odd that DocuComp doesn't directly support it.

### Files with Pictures

While word processing files consist mostly of words, there's the occasional illustration. DocuComp only recognizes the *location* of pictures within a document. It does not compare the pictures themselves. When it encounters a picture, it instead shows the word "[picture]". If there is a difference between the location of pictures in the two documents, it will highlight the word with the correct "inserted" or "deleted" style. This can tell you when a picture is moved, inserted, or deleted, but it won't tell you when a picture has been changed but remains in the same place. Also, the picture themselves do not appear in any of DocuComp's displays or composite documents.

### Spreadsheets

DocuComp can compare any file that has been saved as text, so I ran a spreadsheet in Excel and saved two versions as tab-delimited text (one of the "Options" in the "Save As..." dialog). You won't be surprised to hear that DocuComp easily discovered the single change I made between the two files. Since changes can be displayed in the style of your choice (bold-italic-outlined, if you want), it becomes trivial to find otherwise imperceptible differences between two spreadsheets.

### What DocuComp Compares

DocuComp considers all blanks and tabs to be identical. Thus, a run of 10 spaces is the same as a single space, and tabs also match blanks. It does not compare formatting information at all, except for hard page-breaks, which can optionally be ignored. Thus any font, style, or margin information is simply ignored, and DocuComp will *not* detect changes that purely involve formatting.

DocuComp can optionally ignore or compare footnotes, headers, footers, or any kind of “invisible” text within the document. Since some word processors allow notes within a document that don’t normally print, you can compare just the printable portions of documents, and not worry about wading through all kinds of different internal notes. Although DocuComp won’t compare the content of pictures in your document, it *will* tell you about insertions, deletions, and moves of pictures. These appear in their correct position as the word [picture], in the correct hiliting style. While this is usually adequate, it won’t tell you when a picture itself has changed but remained in the same place.

## The Outer Limits

DocuComp can handle a maximum of 35,000 changes. With this many changes, it needs an 8M Macintosh. The maximum number of changes is lower on Macs with less memory. Since one change can include any number of words, it is unlikely that you will hit this limit in normal use. I didn’t attempt to test this limit.

DocuComp’s default memory allotment in MultiFinder is 512K. More memory makes it go faster, especially for large documents. If you often compare documents of more than about 30 pages, you might want to increase DocuComp’s memory allotment to 1M or more. I didn’t need to do this, since it ran plenty fast with just the 512K allotment.

## Technical Support

I called Advanced Software to report a bug I’d noticed when reading a particular file format. I was quickly connected to Jeff, who seemed to be one of the program’s authors, he knew it so well. When I told him of the problem, he was very interested in how to produce it so he could fix it. He even asked that I send the exact files which were causing the problem (which I did). He then suggested a work-around: saving the file as text-only. I also asked him about the weirdness of scrolling with the arrow-keys, and he acknowledged the problem. The only help here was to refrain from using the arrow keys until they fixed the problem — easy enough to do. He did assure me that it was only a display problem, not a lurking bomb.

The file-reading bug never actually caused the Macintosh to bomb. DocuComp simply displayed an alert informing me of “Fatal error 88” and cleanly quit back to the Finder. The bug turned out to be a problem when MacWrite II saves in the Word 4.0 file format. This is clearly a problem in MacWrite II, since genuine Word 4.0 files do *NOT* have this problem. Also, if you open the MacWrite-II-as-Word-4.0 file in Word 4.0 and then save it, there is no problem.

Advanced Software assured me that they would forward this bug report to Claris, who makes MacWrite II. (The simple fix is to not save files from MacWrite II in Word 4.0 format and then use them directly in DocuComp. Instead, use the Word 3.0 format, which works correctly in MacWrite II, Word 4.0, and DocuComp.)

## Summary

DocuComp works very well. The algorithm it uses for comparisons is very good. It’s fast and works with a variety of document types, including the “text-only” form that every word processor or text editor can produce.

The composite document, showing all deletions, additions, and movements is a very useful representation. If you manually red-line documents and then type in the changes, it also serves as a reference to verify that you’ve got all the changes in place correctly. It even *LOOKS* like the way you’d mark up a document manually. For legal documents, the ability to print line numbers on each page is quite useful. It would also be useful whenever you have any formal document. I’ll be using it on software specifications, design documents, and other such things.

DocuComp has only a few snags, none of which are insurmountable. The arrow-keys scrolling problem is not a big deal, and is acknowledged as “under repair” by the company. The fact that it won’t directly read the original MacWrite format is inconvenient, but not fatal. The problem with reading MacWrite-II-saved-as-Word-4.0 files is clearly a bug in MacWrite II, not DocuComp.

Technical support was very good. People were knowledgeable, friendly, and quick to respond with useful answers. The only down side is that it’s not a toll-free phone call.

If you ever find yourself wondering what the differences between two documents are, then DocuComp is the program for you. If your current method for comparison is staring long hours at the screen, or sitting at your desk with red pen in hand, then DocuComp could bring a tiresome and error-prone chore into the computer age.