

HIDDEN INFORMATION for All to See

By JACK MOLISANI, Senior Member
Orange County Community, Aloha Community

I recently got involved in what, for me, is a new field: computer forensics. Those of you addicted to TV shows on crime scene investigation may know “computer forensics” as the field in which trained technicians search computers for evidence of wrongdoing.

One of the first things I learned in computer forensics is that computers record *incredible* amounts of information about user activity. This information is stored in various places, such as operating system log files, Web browser history files, software application data files, and more. Often the information is hidden and difficult to retrieve, but some can be viewed easily if you know where to look.

It occurred to me that the availability of hidden information has implications beyond trade secret and copyright infringement cases—it could affect the job of any technical communicator. As the saying goes, information is power. Just what kind of information about yourself and your company are you sending out for all the world to see? Shouldn’t you know?

While it takes special forensic tools to access most of the hidden information in computers, some of it is in plain view and can be seen without special tools. This article is about one of the “plain view” instances: information Microsoft *Word* saves about you, your company, and the topic you are writing about, *all of which can be seen by anyone who has access to your document.*

Invisible Ink

In addition to the contents of a document, Microsoft *Word* saves informa-

tion about the document itself. This additional information, called *metadata* (from the Greek *meta*, meaning “higher, beyond”), includes the following:

- Who created the document
- When it was created
- On which machine it was created
- Each party who subsequently opened and saved the document
- If it was saved to a different machine
- Whether the document was saved under a different name
- More

To see a simple example of this retained information, open a Microsoft *Word* file and select **Properties** from the **File** menu. A dialog will appear showing some of this information, such as the document title and author. (See Figure 1.)

Figure 1. Simple metadata stored in Word.

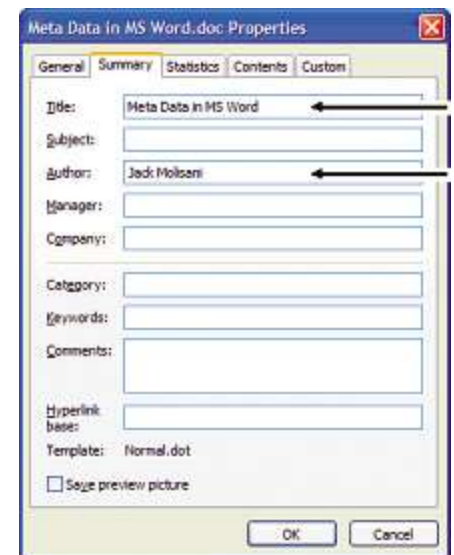


Figure 2. “Recover Text from Any File” reveals more metadata.

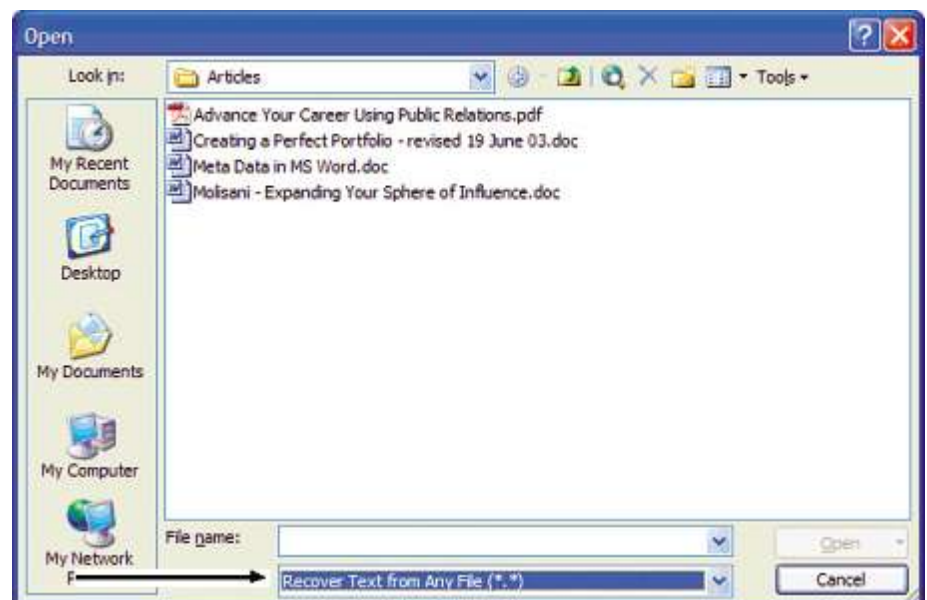
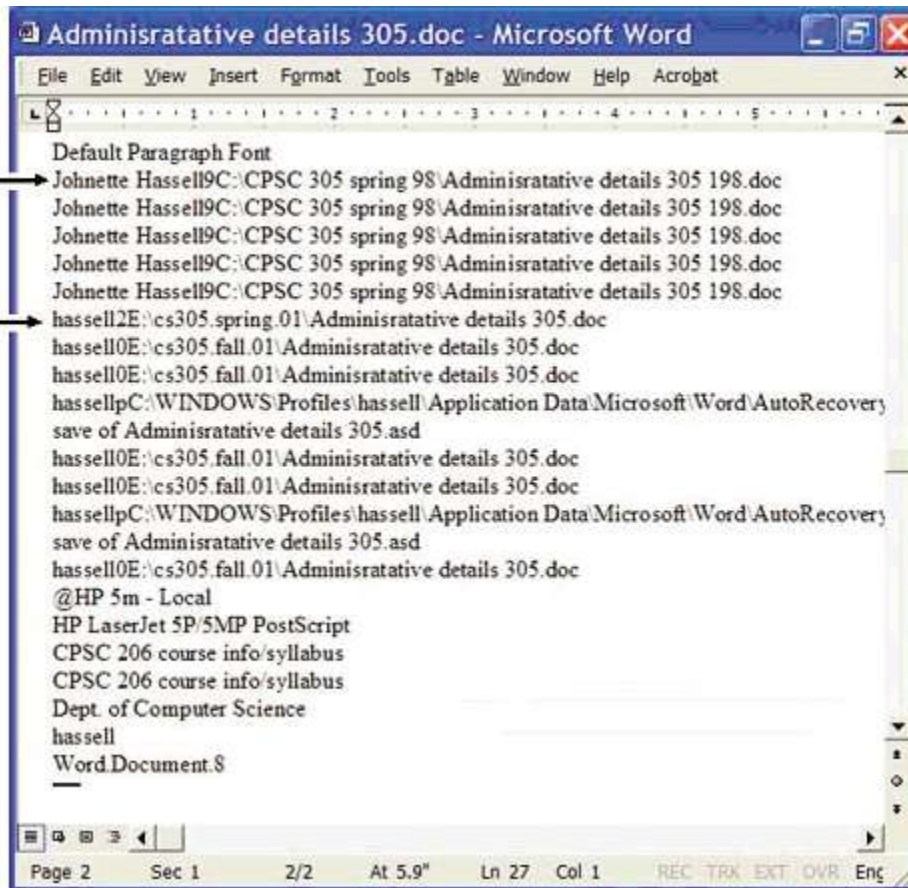


Figure 3. Metadata tracking the movements of a document.



A document written on a corporate PC might display more information, such as the company name, the name of a corporate template being used, etc.

Go ahead and try this and see what metadata is displayed for one of your documents.

Viewing the Hidden Information

To see more metadata stored in a *Word* file, do the following:

1. In Microsoft *Word*, select **File | Open...**
The Open dialog will appear.
2. From the *Files of type* drop-down list, select **Recover Text from Any File (*.*)** and then select and open a *Word* document. (See Figure 2.)
3. When the file opens, page down to the bottom of the file to see metadata such as the following (what you see will vary).

In Figure 3, you can see the original name of the document (“Administra-

Computer Forensics

While my engineering background certainly aided me as a technical writer, it also enabled me to do some interesting side-jobs along the way. One of my professors from Tulane University, Johnette Hassell, started a computer forensics consulting company a few years ago, and recently she had more projects to do than people to do them. When she offered to train me on the latest tools and techniques in computer forensics so I could help her meet some delivery deadlines, I jumped on the opportunity. Somehow I knew this would be more interesting than typing in review comments from subject matter experts... and I was right! If you're interested in learning more about this field, you can see the Web site of Johnette's company at www.ElectronicEvidenceRetrieval.com.

tive details 305 198.doc”) and where it was located (on a machine named “Johnette Hassell9”).

It was then saved under a new name (“Administrative details 305.doc”) in a folder on a different machine (“E:\cs305.fall.01” on “hassell0”):

There is more information that can be viewed, but this exercise shows the type of data Microsoft *Word* stores about your documents.

Should You Care?

While you may not care whether anyone knows how many times you saved a document or the name of the last printer on which it was printed (both are shown in the example), I'll bet you can think of examples of information that could be stored in metadata that you wouldn't want others to have—especially if you share source files with translation companies, resellers, and the like.

Let's look at another example.

A company suspected that an employee was taking home documents that contained trade secrets and selling the information in them to a competitor. The company was granted a court order to image (make an exact bit-by-bit copy of) the hard drive on the employee's home computer, and the company turned the copy over to us for forensic analysis.

While the employee *claimed* he never took documents home, here is what we found:

On the employee's home computer was a document named “How To, Chapter 1.doc”. We opened the document and saw the following in the metadata:

```
Employer\Iran Project\Process
Manual\Section 1.doc
A:\New Manual
Bob Smith7C:\Documents\Iran
Project\User Manual\Section 1.doc
Preferred Customer
New Dell User
C:\Indonesia\How To, Chapter 1.doc
```

Each new piece of information is appended at the bottom of the metadata, so you read the history from the top down. Looking at the list above, you can deduce the following:

(continued on page 43)

(continued from page 21)

1. The employee opened the document “Section 1.doc” on his machine at work. (The name is changed to “Employer” for this article.)
2. He then saved the document to a diskette in drive A.
3. He later saved the document on his home machine under the name “How To, Chapter 1.doc” in a folder named “Indonesia.” (A forensic tool showed that the employee never changed the default name on his home PC; hence, it still showed “Preferred Customer.”)

That’s quite a trail of evidence, huh?

If “Bob Smith” were your employee, you’d be glad to know about metadata. But are there times when you, a law-abiding, loyal employee, wouldn’t want the world to know who had written, edited or changed your document?

Sure. If you’re in negotiations with a customer or vendor, you may not want to draw attention to the fact that you had your legal department touch up certain documents—people sometimes get unnecessarily nervous about such things. Your intra-office politics could be such that you might not want another department to know that your boss had a hand in crafting your memo. Or perhaps you used an external consultant to help you prepare a sales proposal and were not yet ready to disclose that he or she was working on the project.

The point is, by knowing what information is available in your source documents, you can avoid disclosing sensitive or proprietary information (and thus compromising a negotiation, causing an internal or external PR flap, etc.).

How to Protect Yourself and Your Employer

There is not much you can do to keep Microsoft *Word* from storing information in the document metadata, but there *are* things you can do to keep others from seeing it.

The easiest option is just not to send the original *Word* document, but rather to save or print the document as a PDF file and send that. (Metadata is not printed to the PDF file.)

However, if you *must* send the docu-


ment itself (such as when the document must be translated), save it in RTF (rich text format) and send the RTF file. Or first save the document in RTF, convert it back to *Word*, and then send the new *Word* file. Converting a file to RTF saves the formatting but not the metadata.

Caution: Saving a file in RTF strips the document of metadata, but *not* the revision history if you are tracking revisions. Even though *Word* is not displaying revision marks, another contributor to the document may have set revision tracking to “on” but not “visible.” It’s a good practice to check if the document is (or was) recording revisions before saving in RTF to avoid sharing previous revisions—and review comments!

I recommend always converting *Word* documents to PDF, just to be sure.

Epilog

While it is nearly impossible to alter information in a computer to the point where it cannot be found by a competent forensic investigator, you *can* control how much information is made available to recipients of your *Word* documents.

Good luck and good writing! 

When Jack is not saving the world from cyber-crime, he runs ProSpring Technical Staffing (www.prospring.net) and produces Lava-Con: The International Conference on Technical Communication Management, held each September in Honolulu, Hawaii (www.lavacon.org). He can be reached at (310) 831-1929 or at jmolisani@ElectronicEvidenceRetrieval.com.