Taxes redux: You can't open the file until the user tells you to open it

devblogs.microsoft.com/oldnewthing/20090415-00

April 15, 2009

Raymond Chen

One of the so-called taxes of software development on Windows is being respectful of Hierarchical Storage Management. You can't open a file until the user tells you to open it. This rule has consequences for how Explorer extracts information about a file, because what you definitely don't want is for opening a folder full of archived files in Explorer to result in all the files being recalled from tape. (Actually, file recall is just an extreme case of the cost of opening the file. You run into a similar problem if the file is on a slow medium or over a slow network connection. But just to motivate the discussion, I'll continue with the tape scenario.)

What information does Explorer need in order to display a file in a folder? Well, it needs the file name. Fortunately, that can be determined without opening the file, since the file name is how you identify the file in the first place! When you want to open a file, you pass its name. There, you have the name. If you are showing the contents of a folder, you used a function like `FindFirstFile` to get a list of all the files, and the list comes in the form of names. Okay, so the name is easy. (There are still subtleties here, but they are not relevant right now.)

Okay, what's next. The icon. Well, in order to get the icon, we need to know the file type, so let's put the icon on hold for now.

The file creation and modification times can be obtained without opening the file; they come out as part of the `FindFirstFile`, or if you started with a file name, you can recover them with `GetFileAttributesEx`. Either way, you can get that without opening the file. Same goes for the file size.

The other properties, like Title, Author, Summary… They all require that the file be opened, so Explorer disables the property for files that have been archived to tape. You don't want to recall a file from tape just to show its Author in Explorer.

Okay, that leaves just the file type (and the icon, which depends on the file type). Consider the possibilities for how the file type could be determined.

- Read the first few bytes of the file looking for magic numbers.
  Problem: Reading bytes from the file forces a recall.
- Store the type in an alternate data stream.
  Problem: Opening an alternate data stream recalls the file.
- Store the type in the file's metadata.
  Problem: The metadata won't survive being sent as an email attachment or transferred to a file system that doesn't have a place to store that metadata, such as a floppy drive, a CD, or a unix volume.
- Read the first few bytes to determine the type and cache it in the file system.
  Problem: Um, that's just begging the question. We're trying to figure out where in the file system to save it!

What do these problems tell us? The first problem, that reading bytes from a file forces a recall, means that file type information cannot be based on the file contents. The second problem, that reading alternate data streams also force a recall, means that you can't put it in an alternate data stream either. All that's left is storing the type in the metadata.

But the third problem tells us that there isn't much metadata to choose from. Whatever mechanism you use needs to be able to survive being sent as an email attachment or being uploaded to an FTP site. Email attachments in particular are extremely limited. Most email programs, when asked to save an attachment, preserve only one piece of metadata: The filename. (They often don't even preserve the last modified time. And good luck getting them to preserve other ad-hoc metadata.)

All of these problem conspire to rule out all the places you can squirrel away type information, leaving just the filename. It's a sucky choice, but it's the only choice left.

And it means that changing a file's extension means the file type information is destroyed (which, from an end user's point of view, may as well be corruption).

Raymond Chen

Follow