



Home Page

Title Page



Page 1 of 24

Go Back

Full Screen

Close

Quit

Byte Wise

– Internet survival for the rest of us

Martin Vermeer*

December 14, 2005

"Cf. attachment."

I try to open the attachment – no luck. Doesn't open, though it looks like a Microsoft Word file... like some 80% of attachments I have the pleasure to receive.

So I save it to the hard disk and open it. Great... a whole whopping three lines of text, no graphics, no tables, no fancy formatting. Just text. What the heck did the author think, wrapping it into a Word file and not even setting the MIME type right?

*With contributions by Paul Wessel and Jim Luetgert. Thanks!



Home Page

Title Page



Page 2 of 24

Go Back

Full Screen

Close

Quit

Friends keep friends private

This is what I think when someone sends me a group email to 200 people, all of them listed in the Cc: header for everybody else to read.

The space occupied by all these names is only a small thing. Even if the actual message is only five lines. Feel free to waste my disk space folks.

No, much worse is the following scenario: *imagine* if only one of those other recipients were infected by one of those Outlook macro viruses, you know, the ones that read your address book and send themselves forward to everybody in it. Only one virus. Before you know it, you have been “volunteered” by a colleague to be on its address book. And soon on the world’s spammers’ address books too — virus writers and spammers and organized crime are known to work closely together.

This isn’t company you want to find yourself in. But then, you probably are already. Don’t do the same to your friends: *use the BCC: (blind carbon copy) header instead.*



Home Page

Title Page



Page 3 of 24

Go Back

Full Screen

Close

Quit

Even professors do it

If you think that the above things are plain stupid and that reasonably intelligent people would never do a thing like that, think again. I have caught even professors and highly respected colleague scientists committing these sins.

Being very clever in one field doesn't prevent people from doing very stupid things in other contexts.

Remember that personal computing technology, and the Internet, are still so very young. Most current professors became professors long before those happened. And now we are all busy professionals that have no time to spare to engage in acquiring even basic Internet literacy. And it shows. Sigh.

This slide show was composed for the benefit of these busy people.



[Home Page](#)

[Title Page](#)



Page 4 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

A MIME type, any MIME type

Do you know what MIME means? It's one of the standards of the Internet, written up by IETF, the Internet Engineering Task Force, which has done valuable standardization work since day 1 of the Internet.

E-mail is standardized, as is the use of attachments. IETF standards allow for seamless interoperability irrespective of platform choice. Even with Microsoft applications :-)

MIME means "Multipurpose Internet Mail Extensions". Thanks to the MIME type, a mail attachment announces all by itself, what type of file it is and by which application program it can (and may!) be opened.

For example, if we have an MS Word file, the correct MIME type to use would be `application/msword`. If it is an Excel table, use `application/msexcel`. If it is plain text, `text/plain` will do, if HTML ("a Web page"), you should use `text/html`.



Home Page

Title Page



Page 5 of 24

Go Back

Full Screen

Close

Quit

Using MIME the right way

The MIME type is to an e-mail message what — at least under Windows — is for your own documents the *extension*, the three letter ending, of your file names.

If the file name ends in `.doc`, you know — and more importantly, your computer's desk top software knows — that this is a Word document to be opened by MS Word.

If the file name ends in `.xls`, your file is an Excel table to be opened by Excel or equivalent. If it ends in `.pdf`, the proper thing to use is Acrobat Reader.

If the file name ends in `.exe`, you may assume that it is an operating system virus to be run directly :-). And so on.

However, the MIME type is *not the same thing* as the file name extension! A correctly configured mail reader will apply to every attachment that it sends out the correct MIME type — on Windows machines, e.g., by looking at the file name extension.



Home Page

Title Page



Page 6 of 24

Go Back

Full Screen

Close

Quit

...and what about octet-stream?

What then to make of `application/octet-stream`? Well, it says, in effect, *"I don't know"*. More precisely: *"All I know is that this file is made up of bytes, lots of bytes, from start to end. That's all I can say"*. In a MIME compatible mail reader you will not succeed in opening the attachment. You'll have to save it to the hard disk first.

This particular MIME type should only rarely be used. Especially not if you *know* that the attachment is a Word document... because you wrote it yourself in Word.

However, most mail software for Windows will deduce the correct document type in spite of the MIME type being messed up... in Windows, everything is "integrated" with everything else.

If everybody used Windows, MS Office and Outlook, this wouldn't be a problem at all... in the real world it is. That's what IETF standards are for, for *everyone* to follow.



Home Page

Title Page



Page 7 of 24

Go Back

Full Screen

Close

Quit

More MIME soup: RTF

A chapter all of its own are documents in Rich Text Format (RTF).

An RTF document is *not a Word file* — even if you can open and read it in Word. It is a *text file*, which you can open even in a simple text editor like Notepad.

I am all the time receiving RTF documents which have as their MIME type marked `application/msword` instead of the correct `application/rtf`. Don't make that error.

If you are embedded in the matrix of our favourite monopolist, you will never notice anything wrong with this. MS Word opens this stuff just fine, doesn't it? Only if you try to use something else, and have a IETF standards conformant mail reader, there will be wailing and the gnashing of teeth again.



Home Page

Title Page



Page 8 of 24

Go Back

Full Screen

Close

Quit

The multipart/alternative blues

Another wicked but popular pastime is writing your mail message in HTML format. . . or even worse, using suspenders and a belt: both plain-text and HTML. This will only mess up things for your recipient.

The HTML format is unnecessarily powerful for use in e-mails and constitutes an unnecessary security risk — and some junk mail filters will throw your message away unseen.

There really is only one situation in which using HTML in a mail message makes sense: if you want to send out a *link* to something you wrote and have placed on your Web site. This is recommended practice, as the whole document is transferred only to those recipients that *choose* to open it in their browsers.

However, e-mail software is nowadays smart enough to detect, and present in clickable form, also links embedded in a plain text message.

So: *don't use HTML in e-mail. Ever.*



Home Page

Title Page



Page 9 of 24

Go Back

Full Screen

Close

Quit

The Attachment From Hell

Sometimes you may receive an illegible attachment by the name of `winmail.dat`. This can happen *only* if somebody sends you a mail using the Outlook software with the wrong configuration settings. If this happens to you, send your correspondent a *polite* mail pointing out the error of his ways.

If on the other hand somebody sends *you* a complaint, it's *your* job to get your mail software configured properly. How to do that, using information from one of the 60,000-odd Web pages that, according to Google, are dedicated to this fascinating subject, is left as an exercise for the reader.



Home Page

Title Page



Page 10 of 24

Go Back

Full Screen

Close

Quit

Choosing your document type wisely

We saw how important it is to *attach the correct MIME type* to outgoing attachments. If you don't know how to configure your mail software yourself to do that, you owe it to the recipients of your mails to consult with your IT support person. It's minimum politeness.

It is up to *you yourself*, however, to understand that *Microsoft Word is a usually unsuitable document format*. Especially, if your document is meant only to be read and not modified by the recipient. In that case PDF (or even plain text) is the format of choice, and so much simpler. Always looking the same, always secure!

Excellent free or shareware tools for converting your favourite word processing format to PDF exist for your platform too. OpenOffice even offers an "export to PDF" menu option.

...and with PDF, the recipient couldn't even change your text if he wanted to.



Home Page

Title Page



Page 11 of 24

Go Back

Full Screen

Close

Quit

Why using MS Word is (almost always!) a Bad Idea

There are so many reasons... in order of increasing seriousness:

1. Resource wastage
2. Limiting the circle of recipients
3. Unpredictable appearance at receiving end
4. Malware
5. Leaking private information

And so unnecessary! Easy alternatives are readily available, requiring only a little forethought. Plaintext in a simple editor is often the best. Most mail client software comes with a simple editor built-in. Or you can export to text (ASCII) from MS Word or any other word processor.

Or, if your document contains layout elements that would be lost in plaintext, there is PDF, readable by Adobe's Acrobat Reader, which is probably already installed on your system anyway as it is freely downloadable. If it isn't, you should install it and while you're at it, install GhostView as well, another excellent and free viewer for both PDF and PostScript.



Home Page

Title Page



Page 12 of 24

Go Back

Full Screen

Close

Quit

Reason 1: Resource wastage

A Microsoft Word document — like many other word processor document types — is many times bigger on your hard disk than the text it contains would make you believe. Five times, ten times, even twenty times bigger. Some of this information may be legitimately voluminous, like bitmapped illustrations such as photographs, but much of it is plain inefficiency of the storage format.

So you waste network bandwidth capacity, storage capacity on the sending side — if you keep sent mails as most people do — and storage capacity on the receiving side. And... you waste human time on the recipient side, the time it costs to start up a sizable word processing application, when, e.g., a textual message would have shown up directly in the e-mail reader.



Home Page

Title Page



Page 13 of 24

Go Back

Full Screen

Close

Quit

Reason 2: Limiting the circle of recipients

You may have difficulty getting your head around this, but not everybody owns a copy of Microsoft Word, or a Word reader — or even a system that one of these could be installed on. And those that do, often own an older version than you do. It is impolite in the extreme to send people a document in MS Word, especially in a newer format version, without knowing in advance that the recipients (*all* of them) will be able to open it easily.



Home Page

Title Page



Page 14 of 24

Go Back

Full Screen

Close

Quit

Reason 3: Unpredictable appearance at receiving end

A curse well known amongst people in the business of exchanging Word documents. Depending on what version of Word you happen to have, what fonts you have installed and the phase of the moon, the layout changes. Words, sentences and pages break in different places, and not even the number of pages remains constant — something that matters for publishers. And if, like most people, you apply “finger painting” instead of structural editing, the document may arrive completely messed up, e.g., with figures and captions landing on different pages on the recipient’s screen.

If a document is sent out for reading only, PDF is a so much better choice. What you see is what the other guy gets. Well, mostly at least. Fonts can be a headache also with PDF.



Home Page

Title Page



Page 15 of 24

Go Back

Full Screen

Close

Quit

Reason 4: Malware

Microsoft Word contains a programming (macro) language, and therefore can carry macro viruses. The habit of using attachments in “intelligent” formats, and clicking on them thoughtlessly, is also more generally risky, a form of cyber-promiscuity. And the tight integration among the various MS Office elements and with Internet Explorer makes virus propagation even easier.

(You do keep your virus signatures updated, don't you? And you never click on dubious looking attachments, right? And you keep your software updated with the latest security patches too, okay?)



Home Page

Title Page



Page 16 of 24

Go Back

Full Screen

Close

Quit

Reason 5: Leaking private information

Word documents contain lots of “stuff” besides what you put there intentionally. Some of it is harmless, like the huge “empty” areas that go pop when you compress or “zip” the file. Some is not so harmless. Word docs contain lots of “metadata”, like your identity and that of your computer. Not long ago this easily forgotten reality came back to bite British prime minister Tony Blair in the bacon (<http://www.computerbytesman.com/privacy/blair.htm>).

Real life Word documents often contain, e.g., fragments of earlier documents, even such that you thought you had deleted. If you value your privacy — or want to get away with shady dealings, but don't quote me on this — *don't* use Word attachments!

Read also: <http://www.oreillynet.com/pub/wlg/4384>,
<http://news.bbc.co.uk/2/hi/technology/3154479.stm>,
<http://www.ecomputerlaw.com/articles/mhtml/metadata.htm>,
http://www.tvtechnology.com/features/Net-soup/f_fb_leaky_documents.shtml,
http://news.com.com/2100-7344_3-5170073.html,
http://www.forbes.com/technology/2005/12/13/microsoft-word-merck_cx_de_1214word.htm.



Home Page

Title Page



Page 17 of 24

Go Back

Full Screen

Close

Quit

Beyond WYSIWYG

If you are, in good conscience, writing a Word document to be shared with others, *learn at least to use your tool properly*. I am all the time on the receiving end of Word documents that have their sectioning headers “finger painted”, violently boldfaced with section numbers added manually.

You see, the trouble with WYSIWYG (What You See Is What You Get) is that *what you see is all you’ll ever get*.¹

Instead, use *styles*! Sections will number themselves and even go automatically into the table of contents, if you are including one. And use *fields* to number and reference figures and tables etc., so you won’t have to re-type their numbers if you insert a new one in-between.

While MS Word isn’t quite the best thing since sliced bread, it does have these power features worth learning about. Let them work for you!

And the generalist skills acquired using one word processor will remain useful when it is obsoleted by others.

¹...and some things you don’t see, others may nevertheless get, cf. above.



Home Page

Title Page



Page 18 of 24

Go Back

Full Screen

Close

Quit

Avoiding the Irish Famine...

Historians tell us that the Irish Famine of 1845-1850, in which a million people perished, was in no small part due to the cultivation of one strain — asexually reproduced, thus genetically identical — of “lumper” potatoes. When disaster struck, in 1845 in the form of the fungus *Phytophthora infestans* immigrating from North America, it hit with devastating totality.

Somehow the depressing *Windows-Explorer-Outlook-Office* computer viruses and Internet worms situation brings this story to my mind... diversity is the path to survival. Replace any or all of these four by something else, to the extent of your abilities. It's not as if there isn't a choice of excellent alternatives... what about *Linux-Firefox-Evolution-OpenOffice*?



[Home Page](#)

[Title Page](#)



Page 19 of 24

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

...by switching to:

Software	Description	Replaces	Note
Linux	Operating system	Windows	"Cold Turkey"
Firefox	Web browser	Internet Explorer	Web standards compliant
Evolution	E-mail	Outlook/OE	Powerful Outlook lookalike
OpenOffice	Office suite	MS Office	Doc format compatible

The radical thing to do is go Cold Turkey and replace your operating system by Linux... this means replacing all your other software too by Linux equivalents.

If you don't feel quite up to this, try replacing one application at a time. All of these packages are free software, and Firefox and OpenOffice are available for Windows and Macintosh too in addition to Linux. At least Firefox is technically clearly superior to its commercial counterpart and a pleasure to use. And under active development!



Home Page

Title Page



Page 20 of 24

Go Back

Full Screen

Close

Quit

The many cultural circles of computing

Microsoft Office is a very popular software package. However, not everyone has it, or even values having it. For some platforms, it doesn't even exist: if you *really* want to annoy a scientist, send him a Word attachment on his Unix or Linux workstation!

Scientists are one example of a *sub-culture* in computing. Their preferences are the Unix operating system or equivalent, L^AT_EX and its tools for document preparation, and powerful command line tools for routinely doing things the rest of us wouldn't even dream of. But *not MS Office*.

There are other such subcultures. Macintosh users are a special lot too: they include graphic design artists and architects, folks even less computer literate than your average Windows user. (Funnily enough the latest Macintosh operating systems are based on the same code base, Berkeley's Unix dialect, that many scientists use.)

These choices are at least in part personality based and thus, drawing these preferences into question gets personal real fast. There is also an ethical dimension to not wanting to use, e.g., Microsoft software. Just as a civilized person will be culturally sensitive in an international real world context, similar sensitivity is called for when interacting by technological means over the Internet.



Home Page

Title Page



Page 21 of 24

Go Back

Full Screen

Close

Quit

Collaborative authoring challenges

Collaborative authoring is one often occurring circumstance where using a traditional word processor — like MS Word — makes perfect sense. Before doing so, however, it is wise to agree which version of which word processor to use. Otherwise the user of the newest version will force everybody else to upgrade!

An inexpensive alternative to MS Word is OpenOffice. It is reasonably good at reading and exporting fairly complex MS Word documents, in addition to PowerPoint and Excel. It is not perfect, but then, different versions of MS Word aren't perfect among themselves either. An option to consider.

Instead of sending the latest version of your document around by e-mail to everybody, consider “posting” it on your website, and sending out only a *link* to it. This saves bandwidth and hard disk footprint on the receiving end, a win-win situation.



Home Page

Title Page



Page 22 of 24

Go Back

Full Screen

Close

Quit

Filling in forms

Filling in forms is one of those things where there is no really good solution that makes everyone happy.

1. Sending a Word document to be filled in. Horribly wasteful of resources, and you cannot be sure that everyone has Word compatible software, or that it will work even reasonably well. Forms have been known to cause problems for OpenOffice.
2. Better is sending an RTF (Rich Text Format) document. Stand-alone RTF editors exist for your favourite platform too.
3. Using PDF. Unfortunately filling in a PDF form somebody sent you would require the purchase of Adobe Acrobat, again something that may restrict your circle of respondents.
4. Using Web forms (CGI, or Common Gateway Interface). Good solution if you can do it. Many system administrators don't allow you to put CGI scripts on your web site for security reasons. Often, however, a small `mailto` script is provided as standard, allowing you to have the filled-in content of your form mailed back to you.
5. Plain text. What's wrong with plain text? If you are of a coding persuasion, you can even automate the processing of plain text email responses by suitable scripts. Simplest is often best.



Home Page

Title Page



Page 23 of 24

Go Back

Full Screen

Close

Quit

Learning, literacy and community (1)

The unfortunate reality of today is, in spite of the enormous strides that personal computer software has taken, that using computers well is helped enormously by knowing the basics. If you think Windows and Windows software are easy to use, think again. Using Windows and the Internet *securely and responsibly* is *not easy at all*, and impossible without a modicum of computing and Internet literacy.

Just as impossible, in fact, as using the road network safely and efficiently in a society that does not insist on people passing driver's licence tests before being allowed behind the wheel, no matter how safe today's cars are. It's simply not a problem that can be effectively addressed by car design — or road design.

Try to imagine a society where you are allowed to drive without a driver's licence. Where cars are being advertised as being drivable without, or with only very minimal, experience, not requiring any formal training! Would you venture onto the streets in such a society, even as a pedestrian?



Home Page

Title Page



Page 24 of 24

Go Back

Full Screen

Close

Quit

Learning, literacy and community (2)

What “learning to drive” means, is *not* learning to operate a car, but *managing traffic situations*, including challenging and hazardous ones. It’s a social skill set at least as much as a technical one.

Similarly, “learning to use a computer” means *not* learning which button to push to achieve which effect, but how to be connected to the Internet, be part of the information age global community — a social situation again — securely and responsibly.

Better get used to it folks. Computers are getting cleverer all the time, but compared to us, they are still pretty dumb. *You’re* the one that should know better!