

Electronic document distribution

or

Why I can't read Microsoft Word attachments

By Tim Bell (HOD, Computer Science)

23 July 1999

We are entering an exciting era when documents can conveniently be stored and exchanged electronically, with the promise of increased efficiency and reduced costs. However, if done incorrectly, electronic documents can cause frustration and inconvenience. In the last few months a number of people have taken the bold step into the electronic world only to receive a negative response from people including myself. This document presents arguments for why some methods of document exchange are appropriate and efficient, and others are not.

I am not arguing that Word should not be used; in fact this document was prepared in Microsoft Word by choice. I am discussing the method of distribution. This document can be read on any system that has normal access to the World Wide Web. It was prompted when today I received two emails that contained nothing but a Word attachment. The attachments took me about 5 minutes to decode, and turned out to be identical. Furthermore, the information in the documents could easily have been sent as a plain text email.

What's wrong with Microsoft Word attachments?

Microsoft Word is fine; it is its use in disseminating documents that is the problem. A number of university documents have been distributed using Microsoft Word, usually as attachments to email. Word's "Send To" command is beguiling (as one user described it), as it instantly emails the document to many people with very little effort. For many of the recipients, only one click is required to view the document instantly. However, it is not safe to assume this for everyone. To read a Word attachment, I must save it to a file, transfer it to my laptop computer (which may or may not be set up), and open the file. For others in my department, they must additionally leave their office, find a Windows computer, and start it up before they can view the file.

Here is a detailed look at MS Word from the user's perspective.

- Microsoft Word is a program designed for *writing* documents, not for reading them. Its tools are designed for authors, not for readers. When running Word, about 10 of the keys on the keyboard are dedicated to moving around the document; pressing any one of the other 80 or so keys will result in the document being modified. Caution must be exercised to make sure this doesn't happen; just one digit inserted in the wrong place could be disastrous!
- The layout of a Word document can change when viewed on a different computer. Recently a colleague in my department was sent a two-page document that printed as four pages on his computer because two forced page breaks happened to have been moved a little from the original. These problems can be caused by having different versions of fonts, different default paper sizes, or different printers and drivers.
- MS Word exists in different versions that are not necessarily compatible. Sometimes a newer version will translate a document from an older version, possibly losing some formatting; and an older version may not cope with a document from a newer version. I have received Word attachments that have crashed my computer because of version incompatibility; reading a short document cost me 10 minutes to get my computer restarted and cleaned up.

- Word is a proprietary format. Some other manufacturers have produced software that can read and write it, but it is not a widely used and carefully regulated standard that is available on a number of different machines. In particular, a number of departments use Unix systems, and Microsoft do not support this platform.
- It has been argued that a community like the university should agree on an institution-wide standard to enable efficient exchange of documents and data. This overlooks the fact that academic departments are also part of a different community: the Computer Science department is part of the international Computer Science community, in which Unix (in various forms) is a widely used operating system, and Latex is a widely used document production system. As a result of these standards, researchers and teachers freely exchange programs, data and documents. If we abandoned such standards, we would isolate ourselves from this community, defeating the role of the University as an institution with strong international connections. Other departments must also base their choice of computer system on what is widely used in their discipline; for example, in music the Macintosh for some time had the best music processing systems, but a few years ago the Acorn computer became popular amongst musicians because a particularly good system became available for it. The Latex system is generally preferred over Word by people using mathematical symbols, and it is based on a simple programming language. Hence it is popular in Maths and Computer Science departments. For other departments, Microsoft Word running on Microsoft windows will be the most appropriate system.
- One could suggest that academics use separate machines for administration, with their specialist machine being used for research and teaching. However, teaching and research are inextricably mixed with administration. When doing research I might email a colleague some data that has been collected, details of how the research budget is progressing, or text for the next grant application. Likewise, for teaching my interactions with students might be about a snippet of program code, or a warning letter from the AAC.
- If a document is being sent to multiple people who will want printed copies of it, then emailing a Word attachment means that each person will have to spend time printing it out. This will usually take more time, and cost more, than photocopying the document and posting it.
- Word documents can contain viruses, and recently a number of people have been bitten by this. Some people (quite rightly) refuse to open Word attachments from people they don't know.
- Word documents are larger than plain text files. The email that prompted me to write this document was 630 Kbytes — nearly a megabyte! It contained 845 characters of text. This is over 700 times the size it needed to be, and was sent to all Heads.
- In some version of Word, documents can contain text that the author had deleted. The text isn't displayed in Word, but can be viewed using other software. While this text will usually be inaccessible, some authors may prefer that the audience does not have a chance of accessing earlier versions of their documents.

So how should documents be distributed?

The best way of distributing a particular document will depend on its size, content, audience, and the level of formality required. For a particular document, consider each of the following methods:

- **Paper:** particularly suitable if it is a form to be filled in and returned, or if the document needs to be referred to frequently or away from the office.
- **Fax:** for those urgent distributions. Fax machines are mature technology, and the audience automatically receives printed copy.
- **Plain text email:** The best format to guarantee that any recipient can instantly read your email. There are many different email reading systems around (e.g. Outlook, Pegasus Mail, Netscape Mail, Exmh), but even the simplest system will present plain text clearly. If you have typed the text in Word, you can copy it and paste it into the mail sending window. Plain text email does not allow formatting such as italics or different fonts, but if you are simply announcing a meeting time then this won't matter.
- **HTML email:** If the formatting in your email is really important, send it using HTML formatting. Most mail readers can interpret this, and even if they can't, HTML text can be read as plain text with a little effort.
- **Email a reference to a web page:** Put the document on a web page and just send a reference to the URL of the web page. This is particularly suitable if the document is likely to be of archival

interest (e.g. agenda and minutes), and it saves clogging up the mail system with a copy of the document being sent to each recipient.

- **HTML:** Hypertext Markup Language is a system that is primarily used for web pages. It is a simple system that dictates the function of text elements (e.g. Heading) rather than the form (e.g. 14 point bold). The pagination and layout of an HTML document may vary from computer to computer, but it will usually be presented well for the particular system it is being viewed on. HTML documents are prepared using a Web authoring tool, or they can be exported from MS Word or Latex. They are viewed using a Web browser; the most popular are Netscape and Internet Explorer.
- **PDF:** Public Domain Format is the basis of the Adobe Acrobat document exchange system, which was designed specifically to solve the problems that I have highlighted. Acrobat is a sophisticated system designed to publish documents on the Web so that their original pagination, layout, fonts and colour are preserved. Documents can either be “distilled” from electronic systems (including Word and Latex) or “captured” from paper. In either case the document can be indexed, it can be searched, and text can be selected and copied (but not edited without some effort). PDF documents are viewed using Adobe’s “Acroread”, which can be downloaded at no cost from the Internet for many different types of computer and operating system.
- **Word attachments:** Yes, a Word attachment can be appropriate. If I’m authoring a paper with a colleague in Word, then we will exchange attachments. However, this is a one to one relationship between consenting parties!

Word can still be used for preparing HTML and PDF documents. Instead of printing the document it is either exported as HTML, or “printed” to a postscript file which is then processed by Acrobat’s “distill” program to produce PDF.

Conclusion

Am I a Luddite, out to destroy innovative systems? Or one of the anti-Microsoft crowd who feel that Bill Gates’ fortune is built on less than adequate products? Actually, I’m enthusiastic about using technology, but it has to be applied appropriately. In the Department of Computer Science we have an electronic document system running on a private intranet that works extremely well; I can view minutes and agenda of any meeting at any time, from almost any computer, even off campus. Every handout given to every class is available on-line for those who have lost theirs (or happened to have missed the class!) Departmental reports, terms of references for committees, and even paper documents received from the university are available to any member of the department who is authorised to access that particular document.

The system that I have just described has been developed over several years, and a number of lessons have been learned along the way. One of the keys to the success of the system has been the use of non-proprietary formats: HTML (Hypertext Markup Language) and PDF (public domain format). Software for viewing both of these formats is available for almost any computer, and can be freely downloaded from the Internet. Documents are exchanged using a mixture of email and the Web.

If you think that Ned Lud was not as insane as he appeared, and if you find it difficult to cope with your document being reduced to a collection of bits, you may want to consider joining the Lead Pencil Club. See the bibliography below for information about this club, which contains numerous stories from people who have replaced their computer with a pencil, and never been happier. However, for those of us who wish to communicate efficiently, an electronic future beckons.

Bibliography

Minutes of the the [sic] Lead Pencil Club, Bill Henderson, Wainscott, N.Y., c1996.

The trouble with computers, Thomas Landauer, MIT Press, c1995.

Silicon Snake Oil, Clifford Stoll, Doubleday, 1995.