

# O formato MS-Word não é um bom formato para o intercâmbio de documentos.

Ramón Flores

6 de Fevereiro de 2002

## 1 Introdução

Desde há algum tempo resulta excessivamente frequente o envio de mensagens de correio com anexos consistentes em arquivos MS-Word, i.e. com arquivos no formato próprio desse processador de textos. Esta prática, se bem frequente, não é muito recomendável, já que o dito formato não é, do meu ponto de vista, uma boa eleição para transmitir informação textual, tal e como se vê em seguida.

Não se trata neste artigo sobre a idoneidade do programa MS-Word como processador de textos, ainda que eu pessoalmente nem goste dele<sup>1</sup> nem o use, senão da (im)pertinência do formato próprio deste programa para o intercâmbio de documentos. O normal é que um processador de textos possa gravar um mesmo documento em vários formatos<sup>2</sup>, e um @usuári@ inteligente deve escolher qual é o formato mais ajeitado para cada caso.

Este documento pode-se achar em vários formatos na página:

<http://members.tripod.com.br/ramonflores/nom-word/>

## 2 O formato ideal

Um bom formato para a troca de informações textuais deveria ser:

- **Inteligível.** I.e. deveria poder ser lido, e a ser possível mesmo manipulado, em qualquer tipo de computador, rolando qualquer sistema operacional. De maneira a ser entendido por todo o mundo sem problemas.
- **Confiável.** Ou seja, deveria reproduzir fielmente as informações que carrega, nem mais nem menos das que @autor@ deseja enviar, e mantendo a forma dada inicialmente pel@autor@. Ademais não deveria comprometer a seguridade nem do emissor nem do receptor.
- **Ligeiro.** I.e. a razão entre o tamanho ocupado pelo arquivo e as informações contidas deveria manter-se baixa.
- **Resistente a erros.** Ou seja, a perda de alguns poucos bytes não deveria dar lugar a uma perda massiva das informações contidas.

## 3 Carências do formato MS-Word

Infelizmente o formato do MS-Word não cumpre nenhum dos requisitos anteriores. Vejamos porquê:

---

<sup>1</sup>Não sou o único, veja-se por exemplo o artigo de Allin Cottrell Word Processors: Stupid and Inefficient, <http://www.ecn.wfu.edu/~cottrell/wp.html>.

<sup>2</sup>Por exemplo, este artigo foi escrito com o processador de documentos livre LyX <http://www.lyx.org>, o qual permitiu gravá-lo como Texto Simples, em L<sup>A</sup>T<sub>E</sub>X, no seu formato próprio, em HTML, em DVI, em Postscript e em PDF.

### 3.1 O formato MS-Word não tem um alto grau de inteligibilidade.

- **Requer software específico e proprietário.**

Este é um formato proprietário, binário e fechado, pelo qual só o processador de textos MS-Word o entende cabalmente, e para isso nem sempre. Se bem o processador de textos MS-Word está amplamente espalhado, é um aplicativo proprietário, e não especialmente barato, que ademais só tem versões para dois sistemas operacionais, o Windows e o Mac OS. Portanto não todo o mundo o tem instalado no seu computador. A Internet construiu-se procurando que os distintos tipos de computadores puderam comunicar-se entre si, para todo o mundo poder estar a vontade. Enviar arquivos em formatos proprietários pressupondo, sem informação confiável, qual é o software que usa @ recipiendári@ agride esse espírito de procurar a intercomunicação, introduzindo ruído na rede.

- **Há problemas entre versões.**

Mesmo para os usuários de MS-Word, há problemas de compatibilidade entre as diversas versões do aplicativo. Assim utilizar as versões mais modernas impede que os usuários de versões mais antigas podam ler as suas mensagens<sup>3</sup>. Usar versões antigas não soluciona sempre o problema, se bem é prática mais segura, pois ainda assim podem aparecer alterações molestas no formato. Os problemas entre Windows e Mac OS são logicamente maiores.

- **Distintas impressoras = Distintas aparências.**

Ainda que @ recipiendári@ dum ficheiro MS-Word tenha a versão adequada do processador de textos, mesmo assim pode haver problemas. A aparência dum documento MS-Word depende da impressora que tenha o sistema, ou mais bem do driver correspondente. Portanto ao formatar um documento MS-Word, tal vez com uma aparência muito trabalhada para atingir uma impressão dada n@ leitor@, não se pode estar 100% certo de qual vai ser a aparência do dito documento num outro computador com uma impressora diferente.

### 3.2 O formato MS-Word não é confiável.

É bem conhecido que a Microsoft não se tem caracterizado pela elaboração de produtos muito seguros. No caso que nos ocupa isto evidencia-se em dois pontos, as facilidades dadas para criar vírus e a sua falta de discrição.

- **Vírus**

O MS-Word possui um sistema de macros bastante potente e não muito seguro. De maneira que hoje em dia é o vector mais comum de propagação de vírus electrónicos. Enviar arquivos MS-Word habitualmente e sem necessidade é uma maneira muito eficiente de favorecer a propagação de vírus. Se só fossem enviados os ficheiros MS-Word precisos, muitíssimos menos que os actualmente enviados, receber um ficheiro deste tipo não esperado seria já um aviso de perigo, e a difusão dos vírus seria muito mais dificultosa. Realmente quer fomentar o espalhamento de vírus pela rede?

Que um particular envie arquivos MS-Word com vírus não acrescenta a sua estima entre os seus conhecidos. Mas no caso de ser uma instituição a fazê-lo, mesmo a oferece-los nas suas páginas web, dá uma péssima imagem da mesma.

- **Indiscrições**

Os documentos MS-Word incluem mais informações das que aparecem a primeira vista, e que em muitas ocasiões @s autor@s dos documentos não desejariam foram fornecidas. Assim ademais do que se lê ao abrir o documento no processador de textos MS-Word, o documento leva informações sobre o autor e o computador no que foi redigido, e também informações sobre versões antigas do documento, que nalguns casos puderam ser confidenciais. Para ler essa informações chega às

---

<sup>3</sup>Há quem diga que isto não é por acaso, senão procurado pela Microsoft para obrigar a@s usuári@s a actualizar os seus aplicativos. Que não o fariam por necessidade de novas características técnicas, mas si para poder entender os arquivos da gente que si os actualizaram.

vezes com abrir o arquivo com um processador de textos distinto do MS-Word. Outras vezes são necessários procedimentos mais elaborados<sup>4</sup>, mas em todo caso as informações estão no arquivo.

### 3.3 O formato MS-Word é pesado.

Com desmedida frequência enviam-se mensagens num ficheiro MS-Word, que só contêm texto simples, e que pudera ir muito bem numa mensagem de texto normal. Só que no primeiro caso ocupa centenas de KB quando como texto simples ocuparia uns poucos KB. Se para acessar a internet tem que pagar, pense nisto antes de enviar ficheiros Word desnecessariamente. Se é afortunad@ e o acesso resulta-lhe gratuito, seja boa pessoa e pense nos demais.

### 3.4 Resistência aos erros do formato MS-Word.

A combinação processador MS-Word formato MS-Word tem muito pouca resistência aos erros. Uma pequena alteração do arquivo dá lugar a que o processador considere-o corrupto e não permita recuperar nenhuma informação. Mas neste caso o problema não é devido unicamente ao formato mas também ao processador. Assim utilizar outro processador de textos<sup>5</sup>, ou o olhar o ficheiro com outras ferramentas<sup>6</sup>, pode permitir recuperar a maior parte da informação.

## 4 Alternativas

Existem diferentes alternativas ao formato MS-Word, nenhuma delas válidas para todas as ocasiões. Da mesma maneira que não nos vestimos igual para uma boda e para jogar futebol, não devemos utilizar o mesmo formato para diferentes finalidades.

### 4.1 Texto simples

Há muitos casos onde o melhor é utilizarmos texto simples; uma carta comercial ou a um amigo, um artigo num grupo de notícias da USENET, ou numa lista de correios, etc. Sempre que possível deve preferir-se o texto simples, pois é o formato mais inteligível, ligeiro e resistente a erros (sem redundâncias), e 100% confiável. Ademais permite incluir facilmente nas respostas trechos da mensagem contestada. Podemos dormir com smoking, mas é muito mais cómodo fazê-lo com um pijama.

Caso de ser necessário enviar um documento com uma aparência elaborada existem formatos mais adequados que o MS-Word:

### 4.2 RTF

O RTF, acrónimo de Rich Text Format, é um formato que permite uma formatação muito elaborada, sendo propriedade da Microsoft. Mas a diferença do formato MS-Word é aberto, pois as especificações são públicas e facilmente verificáveis ao ser legível pelas pessoas. É por isso que é lido e escrito correctamente por uma grande variedade de processadores de texto. Ademais ao ser legível pelas pessoas, sempre se pode recuperar muita informação mesmo que for levemente danado.

Está pois indicado para quando se quer enviar um documento, que possa ser manipulado pel@ recipiendári@, e cuja forma principal de leitura seja impresso.

#### 4.2.1 Como gravar um arquivo com formato RTF no MW-Word?

Ir ao menu

Arquivo -> Gravar como -> Gravar como tipo:

---

<sup>4</sup>Em Unix por exemplo pode-se usar o comando `strings` para obter texto "oculto" num arquivo MS-Word.

<sup>5</sup>Por exemplo o processador livre Abiword, <http://www.abiword.com>.

<sup>6</sup>Por exemplo o já citado comando `strings`.

e escolher

Formato RTF (\*.rtf)

### 4.3 HTML

O HTML, acrónimo de Hypertext Markup Language, é um formato aberto, público e legível pelas pessoas, pelo qual é entendido por muitos aplicativos informáticos; evidentemente todos os navegadores de internet, muitos processadores de textos, clientes de correio, etc. É um formato pensado para ver documentos na tela, mas não é tão bom para documentos que se vão imprimir. Também não é conveniente abusar deste formato. Algum conhecido cliente de correio envia por defeito mensagens de correio duplas: texto e html, o qual é totalmente desnecessário se o que se quer enviar é uma mensagem de texto.

#### 4.3.1 Como gravar um arquivo com formato HTML no MS-Word?

Ir ao menu

Arquivo -> Gravar como -> Gravar como tipo:

e escolher

Página Web (\*.htm; \*.html)

Contudo deve sublinhar-se que o MS-Word produz ficheiros html de bastante pouca qualidade, muito pesados e não sempre respeitantes dos estâdares. Neste aspecto é bastante melhor a qualidade dum ficheiro rtf produzido por este processador.

### 4.4 PDF

Acrónimo de Portable Document Format, o PDF é um formato aberto, binário e propriedade de Adobe. Mas as especificações são públicas e bem publicitadas, existindo aplicativos que lêem e escrevem PDF em qualquer sistema operacional. Se o que se deseja é enviar documentos com muito boa aparência, que não dependa da impressora a usar, e que não possa ser alterado facilmente pelo destinatário, este é o formato de eleição. De facto é usado cada vez mais pelas empresas para difundir informação corporativa: catálogos, especificações técnicas, etc. Por outra parte não é muito resistente a erros, ao ser binário, e também não é especialmente ligeiro.

O MS-Word não permite gravar em PDF, mas há formas de fazê-lo com software livre. Uma completa descrição de como o fazer pode achar-se na página web Using Ghostscript to Make PDF Files, <http://www.sp.uconn.edu/~fdb94001/using-gs/make-pdf.html>.

## 5 Quando é apropriado empregar o formato MS-Word

Dizia anteriormente que não há nenhum formato apropriado em todas as circunstâncias, senão que segundo a situação haverá que usar um ou outro. Existe também uma circunstância propícia para empregar o formato MS-Word: quando se sabe positivamente que o destinatário tem o processador de textos MS-Word, uma versão que possa manejar sem problemas o arquivo enviado, e ademais vai ora trabalhar no dito documento, ora empregar partes dele noutros.

## 6 Porquê deste documento

A verdade nunca gostei do processador de textos MS-Word, pelo qual nunca o tive no meu computador. Porém há bastante gente que pressupõe que eu tenho o dito processador, e envia-me arquivos nesse formato. E já cansei de ver, depois de passar bastante trabalho para os abrir, que na maioria dos casos o formato era totalmente supérfluo.

Há uns poucos meses bati nas minhas navegações pela rede com um artigo de Jeff Goldberg, que partindo duma situação semelhante a minha tomara-se o trabalho de redigir uma alegação contra o abuso do envio indiscriminado de documentos em formato MS-Word. Pensei em traduzi-lo, mas o fui deixando e há um par de semanas Richard Stallman, o pai do software livre, deu a lume um artigo incitando a rebelar-se contra este abuso MS-Wordiano. Assim que comecei a traduzir, mas já postos abandonei a ideia inicial e redigi um artigo próprio, bastante inspirado isso si no de Jeff Goldberg. Não pretendo portanto ser muito original, tão só contribuir dentro das minhas possibilidades nesta campanha educativa. Pois disso se trata, duma campanha para espalhar boas maneiras na rede.

## 6.1 Páginas relacionadas

- Jeff Goldberg, *MS-Word is Not a document exchange format* <http://www.goldmark.org/netrants/no-word/attach.html>.
- Richard Stallman, *We can put an end to Word attachments* [http://linxtoday.com/news\\_story.php3?ltsn=2002-01-11-002-20-OP](http://linxtoday.com/news_story.php3?ltsn=2002-01-11-002-20-OP).
- Martin Vermeer's, *Plaintext: In praise of practical e-mail hygiene*, <http://www.netby.dk/Oest/Europa-Alle/vermeer/plain.html>.
- Michael A. Gurski, *Why No Word?*, [http://userpages.umbc.edu/~mgurski/rants/why\\_no\\_word.html](http://userpages.umbc.edu/~mgurski/rants/why_no_word.html).
- Chris Mason, *Why no Word?* [http://www.unixzone.com/why\\_no\\_word.html](http://www.unixzone.com/why_no_word.html).
- Stéphane Bortzmeyer, *After Word: l'avenir du traitement de texte*, <http://www.internatif.org/bortzmeyer/afterword/afterword.html>.
- Juhapekka Tolvanen, *Miksi on typerää postittaa sähköpostin liitetiedostona MS Word -dokumenteja*, <http://www.cc.jyu.fi/~juhtolv/mswordmail.html>.
- Sam Steingold, *No Proprietary Binary Data Formats*, <http://www.podval.org/~sds/data.html>.
- Adam Barr, *Proposal: The Open Data Format Initiative*, <http://www.osopinion.com/perl/story/16034.html>.
- Scott Mcneal, *The Case for Open File Formats*, [http://www.computerworld.com/cwi/community/story/0,3201,NAV65-1797\\_STO64689,00.html](http://www.computerworld.com/cwi/community/story/0,3201,NAV65-1797_STO64689,00.html).
- Kevin Reichard, *Editor's Note: The Importance of Open File Formats*, <http://www.linuxplanet.com/linuxplanet/reviews/1934/1/>.